PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :		(11) International Publication Number:	WO 92/13951
C12N 15/14, 15/81	A1	(43) International Publication Date:	20 August 1992 (20.08.92)

(21) International Application Number: PCT/US92/01015

(22) International Filing Date: 4 February 1992 (04.02.92)

(30) Priority data:
650,040 4 February 1991 (04.02.91) US

(71) Applicant: THE SALK INSTITUTE BIOTECHNOLOGY/INDUSTRIAL ASSOCIATES, INC. [US/US]; 505 Coast Boulevard South, La Jolla, CA 92138 (US).

(72) Inventors: DAVIS, Geneva, Ruth ; 3083 East Fox Run Way, San Diego, CA 92111 (US). PROVOW, Sally, Ann ; 11005 Sagittarius Road, San Diego, CA 92126 (US).

(74) Agents: WATT, Phillip, H. et al.; Fitch, Even, Tabin & Flannery, Room 900, 135 South LaSalle Street, Chicago, IL 60603 (US).

(81) Designated State: JP.

Published
With international search report.

(54) Title: PRODUCTION OF HUMAN SERUM ALBUMIN IN METHYLOTROPHIC YEAST CELLS

(57) Abstract

Human serum albumin, a naturally occurring single chain polypeptide, is prepared by growing methylotrophic yeast transformants containing in their genome at least one copy of a DNA sequence operably encoding HSA, in operational association with a DNA sequence encoding a secretion signal sequence selected from: (1) the S. cerevisiae alpha mating factor pre-pro sequence (including the proteolytic processing site: lys-arg), or, (2) the native HSA secretion signal sequence, wherein both the HSA encoding fragment and the secretion signal sequence are maintained under the regulation of a promoter region of a gene of a methylotrophic yeast, under conditions allowing expression of said DNA sequences, and secretion of HSA into the culture medium. Also disclosed are novel DNA fragments and novel recombinant yeast strains which are useful in the practice of the present invention.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FI	Finland	ML	Mali
AU	Australia	FR	France	MN	Mongolia
BB	Barbados	GA	Gabon	MR	Mauritania
BE	Belgium	GB	United Kingdom	MW	Malawi
BF	Burkina Faso	GN	Guinea	NL	Netherlands
BG	Bulgaria	GR	Greece	NO	Norway
BJ	Benin	HU	Hungary	PL	Poland
BR	Brazil	IE	Ireland	RO	Romania
CA	Canada	IT	Italy	RU	Russian Federation
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland			SN	Senegal
CI	Côte d'Ivoire	KR	Republic of Korea	SU	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
CS	Czechoslovakia	LK	Sri Lanka	TG	Togo
DE	Germany	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		
ES	Spain	MG	Madagascar		

PRODUCTION OF HUMAN SERUM ALBUMIN IN METHYLOTROPHIC YEAST CELLS

Field of the Invention

This invention relates to a process of recombinant DNA technology for producing human serum albumin (HSA) peptides in methylotrophic yeast such as Pichia pastoris. Methylotrophic yeast transformants containing in their genome at least one copy of a DNA sequence operably encoding an HSA peptide under the regulation of a promoter region of a gene of a methylotrophic yeast and the S. cerevisiae alpha-mating factor (AMF) pre-pro sequence are cultured under conditions allowing the expression of HSA peptides into the culture medium. The invention further relates to the methylotrophic yeast transformants, DNA fragments and expression vectors used for their production and cultures containing same.

Background of the Invention

Human serum albumin (HSA) is a naturally-occurring, single-chain polypeptide, and is the most abundant protein found in the plasma of adult humans. The concentration of albumin circulating throughout the human body is about 40 mg/ml, corresponding to about 160 g of albumin for a 70 Kg adult male. This protein aids in the body's maintenance of osmotic pressure, and functions in the binding and transport of a wide variety of species, e.g., copper, nickel, calcium, bilirubin, protoporphyrin, long chain fatty acids, prostaglandins, steroid hormones, thyroxine, triiodothyronine, cystine, and glutathione. One particularly large scale application of albumin is the administration of albumin to patients with circulatory failure or with albumin depletion. For example, it is reported that in excess of

10,000 kilograms of purified albumin are administered annually in the United States alone for such purpose.

Naturally occurring HSA contains 585 amino acids. The amino acid sequence of HSA, as reported in the literature, is as set forth in Sequence ID No. 1, provided below.

The amino acid sequence of HSA has been disclosed, for example, by Meloun et al., in FEBS Lett. 58: 134-137 (1975); Lawn et al., in Nucleic Acids Research 9:6103-6114 (1981); and Dugaiczyk et al., in Proc. Natl. Acad. Sci. USA 79: 71-75 (1982).

The molecule in natural form contains 17 disulfide linkages and arises from an about 609 amino acid precursor molecule containing an 18 amino acid pre-peptide, a six amino acid pro-sequence, and the 585 amino acid mature molecule. (see, e.g., Dugaiczyk, et al., supra).

Since isolation of HSA from natural sources is technically difficult, expensive, and time consuming, recent efforts have centered on the development of efficient recombinant methods for the production of HSA.

Of the hosts widely used for the production of heterologous proteins, probably E. coli and Saccharomyces cerevisiae (Baker's yeast) are the best understood. However, E. coli tends to produce HSA in an insoluble, aggregate form. In addition, since E. coli produces endotoxins, the recombinantly expressed product must be subjected to extensive purification treatment to ensure its safety for use in human subjects.

Yeasts can offer clear advantages over bacteria in the production of heterologous proteins. Such advantages include the ability of yeast to secrete heterologous proteins into the culture medium. Secretion of proteins from cells is generally superior to production of proteins in the cytoplasm. Secreted products are obtained in a higher degree of initial

purity and their further purification is easier to accomplish without cellular debris. In the case of sulfhydryl-rich proteins, there is another compelling reason for the development of hosts capable of secreting them into the culture medium: their correct tertiary structure is produced and maintained via disulfide bonds.

The secretory pathway of the cell and the extracellular medium are oxidizing environments which can support disulfide bond formation [Smith, et al., Science, 229, 1219 (1985)]. In contrast, the cytoplasm is a reducing environment in which disulfide bonds cannot form. Upon cell breakage, too rapid formation of disulfide linkages can result in random disulfide bond formation. Consequently, production of sulfhydryl rich proteins, such as HSA, containing appropriately formed disulfide bonds, can be best achieved by transit through the secretory pathway.

Secretion of authentic biologically active human serum albumin from S. cerevisiae is disclosed in Biotechnology 4: 726-730 (1986) by Etcheverry et al. (employing native HSA signal sequence); Biotechnology 8: 42-46 (1990) by Sleep et al., (employing five different leader sequences: S. cerevisiae alpha mating factor signal sequence, native HSA signal sequence, Kluyveromyces lactis killer signal sequence, a fusion of natural HSA/alpha mating factor sequences, and a fusion of K. lactis killer/alpha mating factor sequences); Nucleic Acids Res. 18: 1308 (1990) by Cousens et al; and Nucleic Acids Res. 18: 6075-6081 (1990) by Kalman et al., (employing S. cerevisiae PH05 signal sequence). In what appears to be the best experiments from among all of these references, HSA is produced in S. cerevisiae by means of an expression cassette containing a DNA sequence encoding mature HSA joined to sequences encoding the Kluyveromyces lactis killer signal sequence fused to the S. cerevisiae alpha mating

factor signal sequence, or the native HSA signal sequence fused to the S. cerevisiae alpha mating factor signal sequence. HSA was secreted into the shake flask culture medium in a concentration up to only about 55 mg/l. In view of the problems usually encountered with up-scaling the production of heterologous proteins in autonomous plasmid-based yeast systems, such as S. cerevisiae, there is no indication that HSA production in S. cerevisiae could be at levels higher than those of the above-described experimental system.

To overcome the major problems associated with S. cerevisiae, e.g. loss of selection for plasmid maintenance and problems concerning plasmid distribution, copy number and stability in fermentors operated at high cell density, a yeast expression system based on the methylotrophic yeast Pichia pastoris has been developed. A key feature making this system unique lies with the promoter employed to drive heterologous gene expression. This promoter, which is derived from the methanol-regulated alcohol oxidase I (AOX1) gene of P. pastoris, is highly expressed and tightly regulated (see e.g. U.S. Pat. No. 4,855,231, issued August 8, 1989). Another key feature of the P. pastoris expression system is the stable integration of expression cassettes into the P. pastoris genome, thus significantly decreasing the chance of vector loss (see U.S. Pat. No. 4,882,279, issued November 21, 1989).

Cytoplasmic production of authentic biologically active human serum albumin from P. pastoris is disclosed in European Patent Application No. 89107459.3, published December 6, 1989 (No. 0 344 459);. The cited document contains very little detail as to the level of production or the purity of HSA obtained. In what appears to be the best experiment, HSA is produced in P. pastoris by means of an expression cassette containing a cloned cDNA sequence encoding mature HSA

under the control of the AOX1 promoter. HSA was obtained from cells grown in shake flask culture (after cell breakage and centrifugation to remove cell debris) at concentrations up to only about 88,000 ng/ml. In view of the problems encountered with the production of HSA in the above-described P. pastoris yeast expression system, there is no indication that HSA production in P. pastoris could be at levels higher than those of the above-described experimental system, or that HSA could be secreted from P. pastoris.

Although P. pastoris has been used successfully for the production of various heterologous proteins, e.g., hepatitis B surface antigen [see Cregg et al., Bio/Technology 5, 479 (1987); and U.S. pat. No. 4,895,800, issued January 23, 1990], lysozyme and invertase [Digan et al., Developments in Industrial Microbiology 29, 59 (1988); Tschopp et al., Bio/Technology 5, 1305 (1987)], endeavors to produce other heterologous gene products in Pichia, especially by secretion, have given mixed results. At our present level of understanding of the P. pastoris expression system, it is unpredictable whether a given gene can be expressed to an appreciable level in this yeast or whether Pichia will tolerate the presence of the recombinant gene product in its cells. Further, it is especially difficult to foresee if a particular protein will be secreted by P. pastoris, and if it is, at what efficiency. Even for S. cerevisiae, which has been considerably more extensively studied than P. pastoris, the mechanism of protein secretion is not well defined and understood.

Summary of the Invention

The present invention provides an expression system suitable for the high level production of HSA. In addition, the present invention provides a powerful

method for the production of secreted HSA peptides in methylotrophic yeast such as Pichia pastoris, which method can be easily scaled up from shake-flask cultures to large fermentors with no loss in productivity and without making major changes in the fermentation conditions. The presently preferred yeast species for use in the practice of the present invention is Pichia pastoris, a known industrial yeast strain that is capable of utilizing methanol as the sole carbon and energy source (methylotroph).

We have surprisingly found that HSA peptides can very efficiently be produced in, and secreted from, methylotrophic yeast such as P. pastoris, by transforming a methylotrophic yeast with, and preferably integrating into the yeast genome, at least one copy of a first DNA sequence operably encoding an HSA peptide, wherein said first DNA sequence is operably associated with a second DNA sequence encoding a secretion signal sequence selected from the S. cerevisiae alpha-mating factor (AMF) pre-pro sequence (including the proteolytic processing site: lys-arg), or the HSA signal sequence, and wherein both of said DNA sequences are under the regulation of a methanol responsive promoter region of a gene of a methylotrophic yeast. Methylotrophic yeast cells containing in their genome at least one copy of these DNA sequences efficiently produce biologically active HSA peptides as a medium secreted product.

Accordingly, this invention relates to a methylotrophic yeast cell containing in its genome at least one copy of a DNA sequence operably encoding an HSA peptide, operably associated with a DNA sequence encoding a secretion signal sequence selected from the S. cerevisiae AMF pre-pro sequence (including the proteolytic processing site: lys-arg), or the HSA signal sequence, wherein both the coding sequence and signal

sequence are maintained under the regulation of a promoter region of a gene of a methylotrophic yeast.

According to another aspect, this invention relates to a DNA fragment containing at least one copy of an expression cassette comprising in the reading frame direction of transcription, the following DNA sequences:

(i) a promoter region of a methanol responsive gene of a methylotrophic yeast,

(ii) a DNA sequence encoding a polypeptide consisting of:

(a) a secretion signal sequence selected from:

(1) the S. cerevisiae AMF pre-pro sequence, including the proteolytic processing site: lys-arg, or

(2) the native HSA secretion signal sequence, and

(b) a DNA sequence encoding an HSA peptide; and

(iii) a transcription terminator functional in a methylotrophic yeast, wherein said DNA sequences are operationally associated with one another for transcription of the sequences encoding said polypeptide.

The DNA fragment according to the invention can be transformed into the methylotrophic yeast cells as a linear fragment flanked by DNA sequences having sufficient homology with a target gene to effect integration of said DNA fragment therein. In this case integration takes place by replacement at the site of the target gene. Alternatively, the DNA fragment can be part of a circular plasmid, which may be linearized to facilitate integration, and will integrate by addition at a site of homology between the host and the plasmid sequence.

The invention further concerns an expression vector containing at least one copy of an expression cassette as described hereinabove.

According to a still further embodiment, the invention relates to a process for producing HSA peptides by growing methylotrophic yeast transformants containing in their genome at least one copy of a DNA sequence operably encoding an HSA peptide, operably associated with DNA encoding a secretion signal sequence selected from the S. cerevisiae AMF pre-pro sequence, or the HSA signal sequence, wherein both the coding sequence and the signal sequence are maintained under the regulation of a promoter region of a methanol responsive gene of a methylotrophic yeast, under conditions allowing the expression of said DNA sequence in said transformants and secreting mature HSA peptides into the culture medium. Cultures of viable methylotrophic yeast cells capable of producing HSA peptides are also within the scope of the invention.

The polypeptide product is secreted into the culture medium at surprisingly high concentrations; the level of secretion of HSA peptides is at least an order of magnitude higher than the best results published in the literature. These present, excellent results are due, in part, to the surprising fact that the S. cerevisiae AMF pre-pro sequence and the HSA signal sequence function unexpectedly well to direct secretion of HSA peptides in methylotrophic yeast such as P. pastoris.

The present invention is directed to the above aspects and all associated methods and means for accomplishing such. For example, the invention includes the technology requisite to suitable growth of the methylotrophic yeast host cells, fermentation, and isolation and purification of the HSA gene product.

P. pastoris is described herein as a model system for the use of methylotrophic yeast hosts. Other useful methylotrophic yeasts can be taken from four genera, namely Candida, Hanensula, Pichia and Torulopsis. Equivalent species from them may be used as hosts herein primarily based upon their demonstrated characterization of being supportable for growth and exploitation on methanol as a single carbon nutriment source. See, for example, Gleeson et al., Yeast 4, 1 (1988).

Brief Description of the Drawings

Figure 1 is a restriction map of plasmid pAO815.

Figure 2 is a restriction map of plasmid pAO856.

Figure 3 is a restriction map of plasmid pHSA111.

Figure 4 is a restriction map of plasmid pHSA211.

Figure 5 is a restriction map of plasmid pHSA212.

Figure 6 is a restriction map of plasmid pHSA214.

Figure 7 is a restriction map of plasmid pHSA216.

In each of the restriction maps provided herein, restriction sites employed for manipulation of DNA fragments, but which are destroyed upon ligation, are indicated by enclosing the notation for the destroyed sites in parenthesis.

For the multi-copy vectors pHSA212, pHSA214 and pHSA216, restriction sites within the "aMF-HSA" construct are not indicated, but are the same as shown in Figure 4 for vector pHSA211. Similarly, the non-HSA encoding BamHI-ClaI fragment in each of vectors pHSA212, pHSA214 and pHSA216 has the same restriciton sites as shown in

Figure 4 for the corresponding BamHI-ClaI fragment of vector pHSA211.

Detailed Description of the Invention

The term "human serum albumin" or "HSA peptide" or simply "HSA", as used throughout the specification and in the claims, refers to a polypeptide product which exhibits similar, in-kind, biological activities to natural human serum albumin (HSA), as measured in recognized bioassays, and has substantially the same amino acid sequence as HSA. It will be understood that polypeptides deficient in one or more amino acids in the amino acid sequence reported in the literature for naturally occurring HSA, or polypeptides containing additional amino acids, or polypeptides in which one or more amino acids in the amino acid sequence of natural HSA are replaced by other amino acids are within the scope of the invention, provided that they exhibit the functional activity of HSA, e.g., aid in the body's maintenance of osmotic pressure, and function in the binding and transport of a wide variety of species, e.g., copper, nickel, calcium, bilirubin, protoporphyrin, long chain fatty acids, prostaglandins, steroid hormones, thyroxine, triiodothyronine, cystine, and glutathione.

The invention is intended to embrace all the allelic variations of HSA. Moreover, derivatives obtained by simple modification of the amino acid sequence of the naturally occurring product, e.g, by way of site-directed mutagenesis or other standard procedures, are included within the scope of the present invention. HSA forms produced by proteolysis of host cells that exhibit similar biological activities to mature, naturally occurring HSA are also encompassed by the present invention.

The amino acids, which occur in the various amino acid sequences referred to in the specification

have their usual, three- and one-letter abbreviations routinely used in the art, i.e.:

Amino Acid	Abbreviation	
L-Alanine	Ala	A
L-Arginine	Arg	R
L-Asparagine	Asn	N
L-Aspartic acid	Asp	D
L-Cysteine	Cys	C
L-Glutamine	Gln	Q
L-Glutamic Acid	Glu	E
L-Glycine	Gly	G
L-Histidine	His	H
L-Isoleucine	Ile	I
L-Leucine	Leu	L
L-Lysine	Lys	K
L-Methionine	Met	M
L-Phenylalanine	Phe	F
L-Proline	Pro	P
L-Serine	Ser	S
L-Threonine	Thr	T
L-Tryptophan	Trp	W
L-Tyrosine	Tyr	Y
L-Valine	Val	V

According to the invention, HSA peptides are produced by methylotrophic yeast cells containing in their genome at least one copy of a DNA sequence operably encoding HSA peptides operably associated with DNA encoding a secretion signal sequence selected from (1) the S. cerevisiae AMF pre-pro sequence (including the proteolytic processing site: lys-arg), or (2) the native HSA secretion signal sequence, both under the regulation of a promoter region of a methanol responsive gene of a methylotrophic yeast.

The term "a DNA sequence operably encoding HSA peptides" as used herein includes DNA sequences encoding the 585 amino acid form of HSA or any other "HSA peptide" as defined hereinabove. DNA sequences encoding HSA are known in the art. They may be obtained by chemical synthesis, or by transcription of messenger RNA (mRNA) corresponding to HSA into complementary DNA (cDNA) and converting the latter into a double stranded cDNA. The

mRNA can be isolated for example, from human liver (see Dugaiczyk, et al., supra). Chemical synthesis of a gene for HSA is, for example, disclosed by Kalman et al., Supra. The requisite DNA sequence can also be removed, for example, by restriction enzyme digest of known vectors harboring the HSA gene. Examples of such vectors and the means for their preparation can be taken from the following publications: Etcheverry et al., supra; Sleep et al., supra; Cousens et al, supra; and Kalman et al., supra, etc. The structure of a preferred HSA gene used in accordance with the present invention is further elucidated in the examples.

The presently preferred promoter region employed to drive the HSA gene expression is derived from a methanol-regulated alcohol oxidase gene of P. pastoris. P. pastoris is known to contain two functional alcohol oxidase genes: alcohol oxidase I (AOX1) and alcohol oxidase II (AOX2) genes. The coding portions of the two AOX genes are closely homologous at both the DNA and the predicted amino acid sequence levels and share common restriction sites. The proteins expressed from the two genes have similar enzymatic properties, but the promoter of the AOX1 gene is more efficient and highly expressed. Therefore, the use of the AOX1 gene is preferred for HSA expression. The AOX1 gene, including its promoter, has been isolated and thoroughly characterized [Ellis et al., Mol. Cell. Biol. 5, 1111 (1985)].

The expression cassette used for transforming methylotrophic yeast cells contains, in addition to a methanol responsive promoter of a methylotrophic yeast gene and the HSA encoding DNA sequence (HSA gene), a DNA sequence encoding a secretion signal sequence selected from the in-reading frame S. cerevisiae AMF pre-pro sequence, including a DNA sequence encoding the processing site: lys-arg (also referred to as the lys-arg encoding sequence), or the native HSA secretion

signal sequence; and a transcription terminator functional in a methylotrophic yeast.

The S. cerevisiae alpha-mating factor is a 13-residue peptide, secreted by cells of the "alpha" mating type, that acts on cells of the opposite "a" mating type to promote efficient conjugation between the two cell types and thereby formation of "a-alpha" diploid cells [Thorner et al., The Molecular Biology the Yeast Saccharomyces, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 143 (1981)]. The AMF pre-pro sequence is a leader sequence contained in the AMF precursor molecule, and includes the lys-arg encoding sequence which is necessary for proteolytic processing and secretion (see e.g. Brake et al., Supra). The AMF pre-pro sequence, including the lys-arg encoding sequence is a 255 bp fragment which is set forth as Sequence ID No. 2.

The native HSA secretion signal sequence is a 24 amino acid fragment, comprising an 18 amino acid pre-peptide, and a six amino acid pro-sequence. These sequences have been described by Dugaiczyk, et al., supra. The nucleotide sequence employed in the practice of the present invention encodes the native HSA secretion signal sequence, employing yeast-preferred codons. See Sequence ID No. 3 for the specific sequence employed herein. Those of skill in the art recognize that numerous other nucleotide sequences will also encode the desired 24 amino acid pre-pro sequence. Such sequences could readily be used instead of the specific sequence set forth in Sequence ID No. 3.

The transcription terminator functional in a methylotrophic yeast used in accordance with the present invention has a subsegment which encodes a polydenylation signal and polydenylation site in the transcript and/or a subsegment which provides a transcription termination signal for transcription from the promoter used in the expression cassette according to the invention (the term

"expression cassette" as used herein and throughout the specification and claims refers to a DNA sequence which includes sequences functional for the expression and secretion process). The entire transcription terminator is taken from a protein-encoding gene, which may be the same or different from the gene which is the source of the promoter used according to the invention.

For the practice of the present invention it is preferred that multiple copies of the above-described expression cassettes be contained on one DNA fragment, preferably in a head-to-tail orientation. It is particularly preferred that four or more copies of the above-described expression cassette be contained on one DNA fragment.

The DNA fragments according to the invention optionally further comprise a selectable marker gene. For this purpose, any selectable marker gene functional in methylotrophic yeast such as P. pastoris may be employed, i.e., any gene which confers a phenotype upon methylotrophic yeast cells, thereby allowing them to be identified and selectively grown from among a vast majority of untransformed cells. Suitable selectable marker genes include, for example, selectable marker systems composed of an auxotrophic mutant P. pastoris host strain and a wild type biosynthetic gene which complements the host's defect. For transformation of $his4^-$ P. pastoris strains, for example, the S. cerevisiae or P. pastoris HIS4 gene, or for transformation of $arg4^-$ mutants, the S. cerevisiae ARG4 gene or the P. pastoris ARG4 gene, may be employed.

If the yeast host is transformed with a linear DNA fragment containing the HSA gene under the regulation of a promoter region of a P. pastoris gene and sequences necessary for processing and secretion, the expression cassette is integrated into the host genome by any of the gene addition or replacement techniques known in the art,

such as by one-step gene replacement [see e.g., Rothstein, Methods Enzymol. 101, 202 (1983); Cregg et al., Bio/Technology 5, 479 (1987)] or by two-step gene replacement methods [see e.g., Scherer and Davis, Proc. Natl. Acad. Sci. USA, 76, 4951 (1979)]. In the gene replacement technique, the linear DNA fragment is directed to the desired locus in the genome of the host organism, i.e., to the target gene to be disrupted, by means of flanking DNA sequences having sufficient homology with the target gene to effect integration of the DNA fragment therein. One-step gene disruptions are usually successful if the DNA to be introduced has as little as 0.2 kb homology with the fragment locus of the target gene; it is however, preferable to maximize the degree of homology for efficiency.

If the DNA fragment according to the invention is contained within or is an expression vector, e.g., a circular plasmid, one or more copies of the plasmid can be integrated at the same or different loci, by addition to the genome, instead of by gene disruption. Linearization of the plasmid by means of a suitable restriction endonuclease facilitates such integration.

The term "expression vector" includes vectors capable of expressing DNA sequences contained therein, where such sequences are in operational association with other sequences capable of effecting their expression, i.e., promoter sequences. In general, expression vectors usually used in recombinant DNA technology are often in the form of "plasmids", i.e., circular, double-stranded DNA loops which, in their vector form, are not bound to the chromosome. In the present specification the terms "vector" and "plasmid" are used interchangeably. However, the invention is intended to include other forms of expression vectors as well, which function equivalently.

In the DNA fragment according to the invention the segments of the expression cassette are "operationally associated" with one another. The DNA sequence encoding HSA peptides is positioned and oriented functionally with respect to the promoter, the DNA sequence encoding the secretion signal sequence, i.e., the S. cerevisiae AMF pre-pro sequence (including the DNA sequence encoding the AMF processing-site: lys-arg) or the HSA signal sequence, and the transcription terminator. Thus, the polypeptide encoding segment is transcribed, under regulation of the promoter region, into a transcript capable of providing, upon translation, the desired polypeptide. Because of the presence of the AMF or HSA signal sequence, the expressed HSA product is found as a secreted entity in the culture medium. Appropriate reading frame positioning and orientation of the various segments of the expression cassette are within the knowledge of persons of ordinary skill in the art; further details are given in the Examples.

The DNA fragment provided by the present invention may include sequences allowing for its replication and selection in bacteria, especially E. coli. In this way, large quantities of the DNA fragment can be produced by replication in bacteria.

Methods of transforming methylotrophic yeast (such as Pichia pastoris) as well as methods applicable for culturing methylotrophic yeast cells containing in their genome a gene for a heterologous protein are known generally in the art.

According to the invention, the expression cassettes are transformed into the cells of a methylotrophic yeast by the spheroplast technique, described by Cregg et al., Mol. Cell. Biol. 5, 3376 (1985). See also U.S. Pat. No. 4,879,231, issued November 7, 1989. Alternatively, the expression cassettes can be transformed into the cells of a

methylotrophic yeast or by the whole-cell lithium chloride yeast transformation system [Ito et al., Agric. Biol. Chem. 48, 341 (1984)], with modification necessary for adaptation to P. pastoris [See U.S. Pat. No. 4,929,555, issued May 29, 1990]. For the purpose of the present invention the spheroplast method is preferred, primarily since it yields a greater number of transformants.

Positive transformants are characterized by Southern blot analysis [Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA (1982)] for the site of DNA integration, Northern blots [Maniatis, Op. Cit., R.S. Zitomer and B.D. Hall, J. Biol. Chem, 251, 6320 (1976)] for methanol-responsive HSA gene expression, and product analysis for the presence of secreted HSA peptides in the growth media.

Transformed strains, which are of the desired phenotype and genotype, are grown in fermentors. For the large-scale production of recombinant DNA-based products in methylotrophic yeast, a three-stage, high cell-density, batch fermentation system is normally employed. In the first, or growth stage, expression hosts are cultured in defined minimal medium with excess glycerol as carbon source. When grown on this carbon source, heterologous gene expression is completely repressed, which allows the generation of cell mass in the absence of heterologous protein expression. Next a short period of glycerol limitation growth is allowed. Subsequent to the glycerol limited growth, methanol is added, initiating the expression of the desired heterologous protein. This third stage is the so-called production stage.

The term "culture" means a propagation of cells in a medium conducive to their growth, and all sub-cultures thereof. The term "subculture" refers to a

culture of cells grown from cells of another culture (source culture), or any subculture of the source culture, regardless of the number of subculturings which have been performed between the subculture of interest and the source culture.

According to a preferred embodiment of the invention, the heterologous protein expression system used for HSA production utilizes the promoter derived from the methanol-regulated AOX1 gene of P. pastoris, which is very efficiently expressed and tightly regulated. This gene can be the source of the transcription terminator as well. The presently preferred expression cassette comprises, operationally associated with one another, a P. pastoris AOX1 promoter, DNA encoding a secretion signal sequence selected from the S. cerevisiae AMF pre-pro sequence (including the DNA sequence encoding the AMF processing site: lys-arg) or the native HSA signal sequence, a DNA sequence encoding mature HSA, and a transcription terminator derived from the P. pastoris AOX1 gene. Preferably, two or more of such expression cassettes are contained on one DNA fragment, in head-to-tail orientation, to yield multiple expression cassettes on a single contiguous DNA fragment.

The presently preferred host cells to be transformed with multiple expression cassettes are P. pastoris cells having at least one mutation that can be complemented with a marker gene present on a transforming DNA fragment. Preferably $his4^-$ (GS115) or $arg4^-$ (GS190) auxotrophic mutant P. pastoris strains are employed.

The fragment containing multiple expression cassettes is inserted into a plasmid containing a marker gene complementing the host's defect. pBR322-based plasmids, e.g., pAO856, are preferred. Insertion of one or multiple copies of the hHSA expression/secretion cassette into parent plasmid pAO856 produces plasmids pHSA111, pHSA211, pHSA212, pHSA214 and pHSA216.

To develop Mut$^-$ expression strains of P. pastoris (Mut refers to the methanol-utilization phenotype), the transforming DNA comprising the expression cassette(s) is preferably integrated into the host genome by a one-step gene replacement technique. The expression vector is digested with an appropriate enzyme to yield a linear DNA fragment with ends homologous to the AOX1 locus by means of the flanking homologous sequences. This approach avoids the problems encountered with S. cerevisiae, wherein expression cassettes must be present on multicopy plasmids to achieve high level of expression. As a result of gene replacement, Mut$^-$ strains are obtained. In Mut$^-$ strains, the AOX1 gene is replaced with the expression cassette(s), thus decreasing the strains ability to utilize methanol. A slow growth rate on methanol is maintained by expression of the AOX2 gene product. The transformants in which the expression cassette has integrated into the AOX1 locus by site-directed recombination can be identified by first screening for the presence of the complementing gene. This is preferably accomplished by growing the cells in a media lacking the complementing gene product and identifying those cells which are able to grow by nature of expression of the complementing gene. Next, the selected cells are screened for their Mut phenotype by growing them in the presence of methanol and monitoring their growth rate.

To develop Mut$^+$ HSA-expressing strains, the fragment comprising one or more expression cassette(s) preferably is integrated into the host genome by transformation of the host with a circular plasmid comprising the expression cassette(s) (wherein the plasmid has been linearized by cutting at a unique restriction site to enhance integration). The integration occurs by addition at a locus or loci having

homology with one or more sequences present on the transformation vector.

Positive transformants are characterized by Southern analysis for the site of DNA integration, by Northern analysis for methanol-responsive HSA gene expression, and by product analysis for the presence of secreted HSA peptides in the growth media. Methylotrophic yeast strains which have integrated one or multiple copies of the expression cassettes at a desired site can be identified by Southern blot analysis. Strains which demonstrate enhanced secretion of HSA may be identified by Northern or product analysis; however, this characteristic is not always easy to detect in shake-flask experiments.

Methylotrophic yeast transformants which are identified to have the desired genotype and phenotype are grown in fermentors. Typically a three-step production process is used. Initially, cells are grown on a repressing carbon source, preferably excess glycerol. In this stage the cell mass is generated in the absence of HSA expression. Next, a short period of glycerol limitation growth is allowed. After exhaustion of glycerol, methanol alone (methanol excess fed-batch mode) or limiting glycerol and methanol (mixed-feed fed-batch mode) are added in the fermentor, resulting in the expression of the HSA gene driven by a methanol responsive promoter. The level of HSA secreted into the media can be determined in a variety of ways, e.g., by Western blot analysis of the media in parallel with an HSA standard, using anti-HSA antisera, or by HPLC after suitable pretreatment of the medium.

The invention is further illustrated by the following non-limiting examples.

Examples

Example 1

The expression vector constructions disclosed in the present application were performed using standard procedures, as described, for example in Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA (1982) and Davis et al., Basic Methods in Molecular Biology, Elsevier Science Publishing, Inc., New York (1986).

All fragments described in this Example were isolated on 0.8% agarose gels. All vectors were digested with the appropriate enzyme then treated with calf alkaline phosphatase. In each case, approximately 10-20 ng of vector were ligated with approximately 100-200 ng of insert. Cells of E. coli strain MC1061 were used as host for transformation of non-M13 vectors; cells of E. coli strain JM103 were used as host for M13 vectors.

The HSA gene was obtained from a pBR322-based plasmid on a 1800 bp HindIII-EcoRI fragment. The HSA encoding fragment employed is set forth in Sequence ID No. 1.

The AMF pre-pro encoding sequence (including the proteolytic processing site: lys-arg) employed in the present study was a 255 nucleotide fragment set forth in Sequence ID No. 2.

This 255 nucleotide fragment was derived from plasmid pAO203. The construction of plasmid pAO203 is described in detail in Example 1f below.

a. Construction of plasmid pHSA211, a one copy vector employing the AMF signal sequence:

Plasmid pHSA211 was constructed as follows:

The synthetic structural gene encoding HSA employed herein, including a codon for methionine at the 5' end, was obtained as an approximately 1800-bp HindIII-EcoRI fragment in plasmid pMET-HSA [Sequence ID No. 1;

see also Kalman, et al., in Nucleic Acids Res. 18:6075 (1990)]. The synthetic gene was prepared with Pichia-preferred codons. Plasmid pMET-HSA was used to transform E. coli strain MC1061. Ampicillin resistant colonies were selected, and DNA from six colonies was analyzed by digestion with HindIII and EcoRI. All six transformants displayed the expected approximately 1800-bp band. One colony was selected for a large-scale plasmid preparation. Subsequently, the entire HindIII-EcoRI insert was sequenced, and was found to agree with that of the original DNA.

The ~1800 bp HindIII-EcoRI fragment from pMET-HSA was isolated and inserted into M13mp10, creating vector pHSA101. Site-directed mutagenesis [Zoller, M.J. and Smith, M. in Methods in Enzymology 100:468 (Wu, R., Grossman, L., and Moldave, K., eds.) Academic Press, New York (1983)] was used to insert EcoRI and BamHI sites at the 3' end of the HSA gene, immediately following the transcription termination codon. The mutagenizing and screening oligonucleotides were of the following sequence:

mutagenizing oligo #1 (SEQ ID No. 4):
TGCTTTGGGTTTGTAAGAATTCGGATCCCGTAATCATGGTCAT
screening oligo #1 (SEQ ID No. 5):
TTGTAAGAATTCGGATCCCGTAAT

Single-stranded vector was prepared from positive plaques and used to transform JM103 cells. The resultant plaques were screened again and single positives were sequenced. One mutagenized clone, pHSA102, was selected and sequenced to verify the addition of EcoRI and BamHI restriction sites.

Plasmid pHSA102 was then used for a second site-directed mutagenesis to remove the StuI site within the gene. Although this StuI site is methylated in DNA prepared from E. coli strain MC1061, and as a result

cannot be cleaved by the enzyme, deletion of this site was performed so that unmethylated DNA from yeast transformants generated with this gene could easily be analyzed by Southern hybridization. Removal of the site was accomplished without alteration of the amino acid sequence; however a HindIII site was created by the change. The mutagenizing and screening oligonucleotides were of the following sequence:

mutagenizing oligo #2 (SEQ ID No. 6):
TGAAAGAGCCTTCAAAGCTTGGGCTGTTGCTAGATT
screening oligo #2 (SEQ ID No. 7):
CTTCAAAGCTTGGGCTG

The resulting mutagenized clone, pHSA103, was sequenced to verify the change.

The approximately 1800-bp HindIII-BamHI fragment from pHSA103 was isolated and inserted into HindIII-BamHI-digested plasmid pAO203. Plasmid pAO203 is comprised of the DNA sequence encoding the αMF pre-pro region (with an EcoRI site at the 5' end), followed by nucleotides which encode the amino acids for processing sites, lys-arg and $(glu\text{-}ala)_2$, which in turn are followed by a HindIII site at the 3' end. Construction of pAO203 is described in Example 1f.

The resulting plasmid, pHSA201, contains, on an EcoRI fragment, DNA encoding the 83-amino acid leader sequence of the αMF prepro region followed by the processing sites, lys-arg and $(glu\text{-}ala)_2$, joined to DNA encoding the HSA gene. The approximately 2050-bp EcoRI fragment was isolated from pHSA201 and cloned into plasmid M13mp18. Transformants were screened with XbaI for the presence of an ~1500-bp band, which indicates that the EcoRI insert is oriented such that the single-stranded template will contain the antisense strand. The resulting vector, pHSA202, was subjected to site-directed

mutagenesis in order to delete the DNA for the $(glu\text{-}ala)_2$ processing sites, the polylinker from the original pMET-HSA plasmid, and the codon for methionine at the 5' end of the gene. The mutagenizing and screening oligos were of the following sequences:

mutagenizing oligo #3 (SEQ ID No. 8):
GTATCTTTGGATAAAAGAGACGCTCACAAGTCTGAAGT
screening oligo #3 (SEQ ID No. 9):
GATAAAAGAGACGCTCAC

The mutagenized plasmid, pHSA203, was sequenced to verify that the gene for HSA was fused directly to the αMF prepro region and lys-arg processing site.

The DNA encoding the prepro αMF-HSA fusion gene was isolated on an approximately 2000-bp EcoRI fragment from pHSA203 and inserted into the unique EcoRI site of Pichia pastoris vector pAO856, described in Example 1j. The ligation was transformed into MC1061 cells and amp^R colonies were selected. Positives were identified by the presence of approximately 3450 and 6500 bp bands upon digestion with SalI. The resulting expression vector, pHSA211 (see Figure 4), comprises one copy of an expression cassette encoding the pre-pro αMF-HSA fusion gene under the control of the Pichia pastoris AOX1 promoter and regulatory regions, as well as the AOX1 transcription termination and polyadenylation signals. In addition, the vector includes the Pichia pastoris HIS4 gene used for selection in His^- hosts and additional 3' AOX1 sequences that can be used to direct integration into the host genome.

The DNA for the entire HSA insert as well as approximately 65 nucleotides each of the promoter and termination regions of pHSA211 were sequenced to verify that no changes had occurred during cloning.

b. Construction of plasmid pHSA111, a one copy vector employing the native HSA signal sequence:

Plasmid pHSA111 is identical to plasmid pHSA211, except pHSA111 contains a signal sequence encoding the native 24-amino acid HSA secretion signal sequence (instead of the AMF secretion signal sequence). The signal sequence was designed, using yeast preferred codons, to (1) place an EcoRI site immediately 5' of the initiation codon, and (2) to delete the codon for methionine from the 5'-end of the HSA gene in vector pMET-HSA. To accomplish these changes, plasmid pHSA103 was subjected to a two-step mutagenesis as previously described. The oligonucleotides employed for mutagenesis were as follows:

mutagenizing oligo #1 (SEQ ID No. 10):
GCT TGC ATG CCT GCA GAA TTC ATG AAG TGG GTT ACT TTC ATT TCT TTG TTG TTC GAC GCT CAC AAG TCT

screening oligo #1 (SEQ ID No. 11):
ACT TTC ATT TCT TTG TTG

mutagenizing oligo #2 (SEQ ID No. 12):
ATT TCT TTG TTG TTC TTG TTC TCT TCT GCT TAC TCT AGA GGT GTT TTC AGA AGA GAC GCT CAC AAG TCT

screening oligo #2 (SEQ ID No. 13):
TCT GCT TAC TCT AGA GGT

The mutagenized clone, pHSA105, was sequenced to verify the addition of the signal sequence and the EcoRI site as well as the deletion of the codon for methionine. The sequence of the final mutagenized gene in pHSA105 comprises the HSA leader sequence set forth in Sequence ID No. 3, fused to the HSA coding sequence set forth in Sequence ID No. 1.

The HSA gene and signal sequence was isolated on an approximately 1830-bp EcoRI fragment from pHSA105 and inserted into the unique EcoRI site of Pichia pastoris vector pAO856 (Example 1j). The resulting

expression vector, pHSA111 (Figure 3), contains one copy of an expression cassette comprising the HSA structural gene and signal sequence under the control of the Pichia pastoris AOX1 promoter and regulatory regions, as well as the AOX1 transcription termination and polyadenylation signals. In addition, the vector includes the Pichia pastoris HIS4 gene used for selection in His⁻ hosts and additional 3' AOX1 sequences that can be used to direct integration into the host genome.

The DNA for the entire HSA insert as well as approximately 65 nucleotides each of the promoter and termination regions of pHSA111 were sequenced to verify that no changes had occurred during cloning.

c. Construction of plasmid pHSA212, a two copy vector employing the AMF signal sequence:

The approximate 3300 bp HSA expression cassette was isolated from pHSA211 on a ClaI-BamHI fragment and inserted back into the ClaI-BglII sites of pHSA211. The ligation was transformed into MC1061 cells and amp^R colonies were selected. Correct transformants demonstrated ~6570 and ~6525 bp ClaI-BamHI fragments upon digestion. The resulting vector, pHSA212 (see Figure 5), contains two copies of the expression cassette linked as tandem-repeat units, as verified by restriction enzyme digests.

d. Construction of plasmid pHSA214, a four copy vector employing the AMF signal sequence:

The approximate 6500 bp ClaI-BamHI fragment from pHSA212, containing two copies of the expression cassette, was isolated and inserted back into the ClaI-BglII sites of pHSA212. The ligation was transformed into MC1061 cells and ampR colonies were selected. Correct transformants demonstrated ~13,000 and 6500 ClaI-BamHI fragments upon digestion. The resulting vector, pHSA214 (see Figure 6), contains four copies of the

expression cassette linked as tandem-repeat units, as verified by restriction enzyme digests.

e. Construction of plasmid pHSA216, a six copy vector employing the AMF signal sequence:

The approximate 6500 bp ClaI-BamHI fragment from pHSA212, containing two copies of the expression cassette, was inserted into the ClaI-BglII sites of pHSA214. The ligation was transformed into MC1061 cells and amp^R colonies were selected. Correct transformants demonstrated ~19,500 and ~6500 ClaI-BamHI fragments upon digestion. The resulting vector, pHSA216 (see Figure 7), contains six copies of the expression cassette linked as tandem-repeat units, as verified by restriction enzyme digests.

f. Construction of plasmid pAO203

The AOX1 transcription terminator was isolated from 20 μg of pPG2.0 [pPG2.0 = BamHI-HindIII fragment of pG4.0 (NRRL 15868) + pBR322] by StuI digestion followed by the addition of 0.2 μg SalI linkers (GGTCGACC). The plasmid was subsequently digested with HindIII and the 350 bp fragment isolated from a 10% acrylamide gel and subcloned into pUC18 (Boehringer Mannheim) digested with HindIII and SalI. The ligation mix was transformed into JM103 cells (that are widely available) and amp^R colonies were selected. The correct construction was verified by HindIII and SalI digestion, which yielded a 350 bp fragment, and was called pA0201.

5 μg of pA0201 was digested with HindIII, filled in using Klenow polymerase, and 0.1 μg of BglII linkers (GAGATCTC) were added. After digestion of the excess BglII linkers, the plasmid was reclosed and transformed into MC1061 cells. Amp^R cells were selected, DNA was prepared, and the correct plasmid was verified by BglII, SalI double digests, yielding a 350 bp fragment, and by a HindIII digest to show loss of HindIII site. This plasmid was called pA0202.

The alpha factor-GRF fusion was isolated as a 360 bp BamHI-PstI partial digest from pYSV201. Plasmid pYSV201 is the EcoRI-BamHI fragment of GRF-E-3 inserted into M13mp18 (New England Biolabs). Plasmid GRF-E-3 is described in EP 206,783. 20 μg of pYSV201 plasmid was digested with BamHI and partially digested with PstI. To this partial digest was added the following oligonucleotides:

5' AATTCGATGAGATTTCCTTCAATTTTTACTGCA 3' (SEQ ID No. 14)

3' GCTACTCTAAAGGAAGTTAAAAATG 5' (SEQ ID No. 15).

Only the antisense strand of the oligonucleotide was kinase labelled so that the oligonucleotides did not polymerize at the 5'-end. After acrylamide gel electrophoresis (10%), the fragment of 385 bp was isolated by electroelution. This EcoRI-BamHI fragment of 385 bp was cloned into pA0202 which had been cut with EcoRI and BamHI. Routinely, 5 ng of vector cut with the appropriate enzymes and treated with calf intestine alkaline phosphatase, was ligated with 50 ng of the insert fragment to produce plasmid pA0203.

g. Construction of plasmid pA0804

Plasmids pA0804 and pA0807 (see Example 1i) are used in the construction of plasmid pA0815.

Plasmid pA0804 has been described in PCT Application No. WO 89/04320. Construction of this plasmid involved the following steps:

Plasmid pBR322 was modified to eliminate the EcoRI site and insert a BglII site into the PvuII site as follows:

pBR322 was digested with EcoRI, the protruding ends were filled in with Klenow Fragment of E. coli DNA polymerase I, and the resulting DNA was recircularized using T4 ligase. The recircularized DNA was used to transform E. coli MC1061 to ampicillin-resistance and transformants were screened for having a plasmid of about 4.37 kpb in size without an EcoRI site. One such

transformant was selected and cultured to yield a plasmid, designated pBR322ΔRI, which is pBR322 with the EcoRI site replaced with the sequence (SEQ ID No. 16):

5'-GAATTAATTC-3'

Plasmid pBR322ΔRI was digested with PvuII, and the linker having the sequence:

5'-CAGATCTG-3'

3'-GTCTAGAC-5'

was ligated to the resulting blunt ends employing T4 ligase. The resulting DNAs were recircularized, also with T4 ligase, and then digested with BglII and again recircularized using T4 ligase to eliminate multiple BglII sites due to ligation of more than one linker to the PvuII-cleaved pBR322ΔRI. The DNAs, treated to eliminate multiple BglII sites, were used to transform E. coli MC1061 to ampicillin-resistance. Transformants were screened for a plasmid of about 4.38 kbp with a BglII site. One such transformant was selected and cultured to yield a plasmid, designated pBR322ΔRIBGL, for further work. Plasmid pBR322ΔRIBGL is the same as pBR322 RI except that pBR322ΔRIBGL has the sequence (SEQ ID No. 17)

5'-CAGCAGATCTGCTG-3'

in place of the PvuII site in pBR322ΔRI.

Plasmid pBR322ΔRIBGL was digested with a SalI and BglII and the large fragment (approximately 2.97 kbp) was isolated. Plasmid pBSAGI5I, which is described in European Patent Application Publication No. 0 226 752, was digested completely with BglII and XhoI and an approximately 850 bp fragment from a region of the P. pastoris AOX1 locus downstream from the AOX1 gene transcription terminator (relative to the direction of transcription from the AOX1 promoter) was isolated. The BglII-XhoI fragment from pBSAGI5I and the approximately 2.97 kbp, SalI-BglII fragment from pBR322ΔRIBGL were combined and subjected to ligation with T4 ligase. The ligation mixture was used to transform E. coli MC1061 to

ampicillin-resistance and transformants were screened for a plasmid of the expected size (approximately 3.8 kbp) with a BglII site. This plasmid was designated pA0801. The overhanging end of the SalI site from the pBR322ΔRIBGL fragment was ligated to the overhanging end of the XhoI site on the 850 bp pBSAGI5I fragment and, in the process, both the SalI site and the XhoI site in pA0801 were eliminated.

Plasmid pBSAGI5I was then digested with ClaI and the approximately 2.0 kbp fragment was isolated. The 2.0 kbp fragment has an approximately 1.0-kbp segment which comprises the P. pastoris AOX1 promoter and transcription initiation site, an approximately 700 bp segment encoding the hepatitis B virus surface antigen ("HBsAg") and an approximately 300 bp segment which comprises the P. pastoris AOX1 gene polyadenylation signal and site-encoding segments and transcription terminator. The HBsAg coding segment of the 2.0 kbp fragment is terminated, at the end adjacent the 1.0 kbp segment with the AOX1 promoter, with an EcoRI site and, at the end adjacent the 300 bp segment with the AOX1 transcription terminator, with a StuI site, and has its subsegment which codes for HBsAg oriented and positioned, with respect to the 1.0 kbp promoter-containing and 300 bp transcription terminator-containing segments, operatively for expression of the HBsAg upon transcription from the AOX1 promoter. The EcoRI site joining the promoter segment to the HBsAg coding segment occurs just upstream (with respect to the direction of transcription from the AOX1 promoter) from the translation initiation signal-encoding triplet of the AOX1 promoter.

For more details on the promoter and terminator segments of the 2.0 kbp, ClaI-site-terminated fragment of pBSAGI5I, see U.S. Pat. No. 4,895,800, European Patent

Application Publication No. 226,846 and Ellis et al., Mol. Cell Biol. 5, 1111 (1985).

Plasmid pA0801 was cut with ClaI and combined for ligation using T4 ligase with the approximately 2.0 kbp ClaI-site-terminated fragment from pBSAGI5I. The ligation mixture was used to transform E. coli MC1061 to ampicillin resistance, and transformants were screened for a plasmid of the expected size (approximately 5.8 kbp) which, on digestion with ClaI and BglII, yielded fragments of about 2.32 kbp (with the origin of replication and ampicillin-resistance gene from pBR322) and about 1.9 kbp, 1.48 kbp, and 100 bp. On digestion with BglII and EcoRI, the plasmid yielded an approximately 2.48 kbp fragment with the 300 bp terminator segment from the AOX1 gene and the HBsAg coding segment, a fragment of about 900 bp containing the segment from upstream of the AOX1 protein encoding segment of the AOX1 gene in the AOX1 locus, and a fragment of about 2.42 kbp containing the origin of replication and ampicillin resistance gene from pBR322 and an approximately 100 bp ClaI-BglII segment of the AOX1 locus (further upstream from the AOX1-encoding segment than the first mentioned 900 bp EcoRI-BglII segment). Such a plasmid had the ClaI fragment from pBSAGI5I in the desired orientation, in the opposite undesired orientation, there would be EcoRI-BglII fragments of about 3.3 kbp, 2.38 kbp and 900 bp.

One of the transformants harboring the desired plasmid, designated pA0802, was selected for further work and was cultured to yield that plasmid. The desired orientation of the ClaI fragment from pBSAGI5I in pA0802 had the AOX1 gene in the AOX1 locus oriented correctly to lead to the correct integration into the P. pastoris genome at the AOX1 locus of linearized plasmid made by cutting at the BglII site at the terminus of the 800 bp

fragment from downstream of the AOX1 gene in the AOX1 locus.

pA0802 was then treated to remove the HBsAg coding segment terminated with an EcoRI site and a StuI site. The plasmid was digested with StuI and a linker of sequence:

5'-GGAATTCC-3'

3'-CCTTAAGG-5'

was ligated to the blunt ends using T4 ligase. The mixture was then treated with EcoRI and again subjected to ligating using T4 ligase. The ligation mixture was then used to transform E. coli MC1061 to ampicillin resistance and transformants were screened for a plasmid of the expected size (5.1 kbp) with EcoRI-BglII fragments of about 1.78 kbp, 900 bp, and 2.42 kbp and BglII-ClaI fragment of about 100 bp, 2.32 kbp, 1.48 kbp, and 1.2 kbp. This plasmid was designated pA0803. A transformant with the desired plasmid was selected for further work and was cultured to yield pA0803.

Plasmid pA0804 was then made from pA0803 by inserting, into the BamHI site from pBR322 in pA0803, an approximately 2.75 kbp BglII fragment from the P. pastoris HIS4 gene. See, e.g., Cregg et al., Mol. Cell. Biol. 5, 3376 (1985) and European Patent Application Publication Nos 180,899 and 188,677. pA0803 was digested with BamHI and combined with the HIS4 gene-containing BglII site-terminated fragment and the mixture subjected to ligation using T4 ligase. The ligation mixture was used to transform E. coli MC1061 to ampicillin-resistance and transformants were screened for a plasmid of the expected size (7.85 kbp), which is cut by SalI. One such transformant was selected for further work, and the plasmid it harbors was designated pA0804.

Plasmid pA0804 has one SalI-ClaI fragment of about 1.5 kbp and another of abut 5.0 kbp and a ClaI-ClaI fragment of 1.3 kbp; this indicates that the direction of

transcription of the HIS4 gene in the plasmid is the same as the direction of transcription of the ampicillin resistance gene and opposite the direction of transcription from the AOX1 promoter.

The orientation of the HIS4 gene in pA0804 is not critical to the function of the plasmid or of its derivatives with cDNA coding segments inserted at the EcoRI site between the AOX1 promoter and terminator segments. Thus, a plasmid with the HIS4 gene in the orientation opposite that of the HIS4 gene in pA0804 would also be effective for use in accordance with the present invention.

h. Construction of plasmid pA0815

Plasmid pA0815 was constructed by mutagenizing plasmid pA0807 to change the ClaI site downstream of the AOX1 transcription terminator in pA0807 to a BamHI site. The oligonucleotide used for mutagenizing pA0807 had the following sequence (SEQ ID No. 18):

5' GAC GTT CGT TTG TGC GGA TCC AAT GCG GTA GTT TAT 3'.

The mutagenized plasmid was called pA0807-Bam. Plasmid pA0804 was digested with BglII and 25 ng of the 2400 bp fragment were ligated to 250 ng of the 5400 bp BglII fragment from BglII-digested pA0807-Bam. The ligation mix was transformed into MC1061 cells and the correct construct was verified by digestion with Pst/BamHI to identify 6100 and 2100 bp sized bands. The correct construct was called pA0815 (See Figure 1).

i. Construction of plasmid pA0807

1. Preparation of f1-ori DNA

f1 bacteriophage DNA (50 μg) was digested with 50 units of Rsa I and Dra I (according to manufacturer's directions) to release the ≈458 bp DNA fragment containing the f1 origin of replication (ori). The digestion mixture was extracted with an equal volume of phenol: chloroform (V/V) followed by extracting the aqueous layer with an equal volume of chloroform and

finally the DNA in the aqueous phase was precipitated by adjusting the NaCl concentration to 0.2M and adding 2.5 volumes of absolute ethanol. The mixture was allowed to stand on ice (4°C) for 10 minutes and the DNA precipitate was collected by centrifugation for 30 minutes at 10,000 x g in a microfuge at 4°C.

The DNA pellet was washed 2 times with 70% aqueous ethanol. The washed pellet was vacuum dried and dissolved in 25 µl of TE buffer. This DNA was electrophoresed on 1.5% agarose gel and the gel portion containing the ≈458 bp f1-ori fragment was excised out and the DNA in the gel was electroeluted onto DE81 (Watman) paper and eluted from the paper in 1M NaCl. The DNA solution was precipitated as detailed above and the DNA precipitate was dissolved in 25 µl of TE buffer (f1-ori fragment).

2. Cloning of f1-ori into Dra I sites of pBR322

pBR322 (2 µg) was partially digested with 2 units Dra I (according to manufacturer's instructions). The reaction was terminated by phenol:chloroform extraction followed by precipitation of DNA as detailed in step 1 above. The DNA pellet was dissolved in 20 µl of TE buffer. About 100 ng of this DNA was ligated with 100 ng of f1-ori fragment (step 1) in 20 µl of ligation buffer by incubating at 14°C for overnight with 1 unit of T4 DNA ligase. The ligation was terminated by heating to 70°C for 10 minutes and then used to transform E. coli strain JM103. Amp^R transformants were pooled and superinfected with helper phage R408. Single stranded phage were isolated from the media and used to reinfect JM103. Amp^R transformants contained pBRf1-ori which contains f1-ori cloned into the Dra I sites (nucleotide positions 3232 and 3251) of pBR322.

3. Construction of plasmid pA0807

pBRf1-ori (10 μg) was digested for 4 hours at 37°C with 10 units each of Pst I and Nde I. The digested DNA was phenol:chloroform extracted, precipitated and dissolved in 25 μl of TE buffer as detailed in step 1 above. This material was electrophoresed on a 1.2% agarose gel and the Nde I - Pst I fragment (approximately 0.8 kb) containing the f1-ori was isolated and dissolved in 20 μl of TE buffer as detailed in step 1 above. About 100 ng of this DNA was mixed with 100 ng of pA0804 that had been digested with Pst I and Nde I and phosphatase-treated. This mixture was ligated in 20 μl of ligation buffer by incubating for overnight at 14°C with 1 unit of T4 DNA ligase. The ligation reaction was terminated by heating at 70°C for 10 minutes. This DNA was used to transform E. coli strain JM103 to obtain pA0807.

j. Construction of plasmid pA0856

Expression vector pA0856 is the same as vector pA0815 except for the following differences:

1. The BamHI site at the 3' end of the transcription termination region in pA0815 is changed to a HindIII-BamHI double site in pA0856. This modification allows each gene in a multi-copy strain, developed using the pA0856 parent vector, to be isolated on a HindIII fragment for independent sequencing during verification.

2. The EcoRV site in the pBR322 region between the transcription terminator and the HIS4 gene has been deleted in pA0856; thus, pA0856 contains a single EcoRV site in the AOX1 3' region. Thus, when a pA0856-based expression vector is integrated into the AOX1 3' region (see #3, below), the entire vector can be recovered on an EcoRV fragment. This will be useful for Southern analyses and strain verification.

3. A NotI site has been added to the 3' AOX1 region in pA0856. A pA0856-based expression vector linearized with NotI can be used for site-directed

integration into the AOX1 3' region. In contrast, multi-copy vectors based on pAO815 can be directed into the HIS4 locus only.

4. The BglII site at the 3' end of the AOX1 3' region has been deleted in pAO856. Multi-copy vectors constructed with pAO815 are generated by isolating the expression cassette on a BglII-BamHI fragment and inserting this fragment back into the BamHI site as a tandem-repeat unit. There are two problems associated with this approach. First, the BglII-BamHI fragment can insert in either orientation; thus, approximately half of the insertions are incorrect and the number of transformants which have to be screened by mini preps and enzyme digestion is increased. Second, inverted-repeat units usually lead to recombination events and, as a result, transformants are more difficult to screen because there is a wide variety of digestion patterns. In pAO856-based vectors, the expression cassette can be isolated on a ClaI-BamHI fragment and inserted into the ClaI-BglII sites. The fragment can insert in only one orientation, thus reducing the number of transformants which need to be screened and increasing the ease of screening.

Additionally, in pAO815-based vectors, the BglII-BamHI digest also liberates an approximately 2400-bp BglII fragment which contains the E. coli ori and amp resistance gene. Any agarose gel-isolated expression cassette of approximately 2400-bp or longer inevitably also contains the 2400-bp BglII fragment (due to the limitations of gel separations). The BglII fragment can self-ligate and transform the host. In fact, because the BglII plasmid is so small, it is a "preferred" transforming fragment, and even a seemingly miniscule amount in a fragment prep leads to a large number of incorrect transformants. In pAO856-based vectors, the

ori and amp gene are contained on a BamHI-ClaI fragment which does not easily self-ligate.

Vector Construction of pA0856

1. Removal of EcoRV site

The approximately 2000-bp EcoRI-SalI fragment from pA0804 (described in Example 1f) containing the AOX1 transcription termination region, 350 bp of pBR322, and the 3' half of the HIS4 gene was inserted into M13mp19. Transformants were screened with HindIII digests. Positive transformants exhibited bands of about 1900 and 7500 bp and were called pHIS102.

Three bases in the EcoRV site located in the pBR322 segment immediately following the AOX1 transcription termination region were deleted in pHIS102 by site-directed mutagenesis. The deletion effected removal of the recognition site. The oligonucleotides were as follows:

Mutagenesis oligo (SEQ ID No. 19):
GGCCTCTTGCGGGATGTCCATTCCGACAGC

Screening oligo (SEQ ID No. 20):
TTGCGGGATGTCCATTCC

The mutagenized plasmid, pHIS103, was sequenced to verify the deletion.

2. Alter HindIII-ClaI-HindIII site to a HindIII-BamHI site

Plasmid pHIS103 was used to change the HindIII-ClaI-HindIII cluster of restriction sites at the 3' end of the AOX1 transcription termination region to HindIII-BamHI sites by site-directed mutagenesis.

Mutagenizing oligo (SEQ ID No. 21):
TTTGTGCAAGCTTATGGATCCGCTTTAATGCGGTAGT

Screening oligo (SEQ ID No. 22):
TTATGGATCCGCTTT

The mutagenized plasmid, pHIS103M, was sequenced to verify the changes. The EcoRI-SalI fragment from

pHIS103M was used in the construction of pAO856, which follows below (see part 4).

3. Insert BamHI site in HIS4 gene and at 3' end of 3' AOX sequences; Insert NotI site in 3' AOX1 sequence

Vector pAO804B is similar to pAO804 except that the HIS4 gene in pAO804B contains a BamHI site which was deleted from the HIS4 gene in pAO804. In order to construct pAO804B, the approximately 2700-bp BglII fragment, which is the HIS4 gene, from pYJ8 (NRRL 8-15889) was inserted into the BamHI site of pAO803 (see Example 1g). Correct transformants displayed approximately 3000- and 5000-bp fragments upon digestion with PvuII.

The approximately 1400-bp BamHI-BglII fragment from pAO804B, comprising the AOX1 3' targeting sequences and the first half of the HIS4 gene, was isolated and inserted into the BamHI site of M13mp19. Transformants with the desired orientation displayed fragments of approximately 1300- and 7400-bp upon double-digestion with EcoRV and SalI.

Single-stranded plasmid pHIS804 was used as template for site-directed mutagenesis to insert a BamHI site at the 3' end of the AOX1 3' region.

Mutagenizing oligo (SEQ ID No. 23):
TTCGAGCTCGGTACCTAAGGATCCTGAGATAAATTTCA
Screening oligo (SEQ ID No. 24): TACCTAAGGATCCTGAG
The mutagenized plasmid, pHIS804M1, was sequenced to verify the change.

Plasmid pHIS804M1 was used as a template for site-directed mutagenesis to insert the sequence for a NotI site into the 3' AOX1 region.
Mutagenizing oligo (SEQ ID No. 25):
ACGTTGTCACTGAAGGCGGCCGCAGTATCTACAAACC
Screening oligo (SEQ ID No. 26):
TGAAGGCGGCCGCAGT

The mutagenized plasmid, pHIS804M2, was sequenced to verify the change.

4. Assembly of pAO856

The approximately 1450-bp BamHI fragment from pHIS804M2, comprising the first half of the HIS4 gene and the 3' AOX1 sequence, was isolated and inserted into the BamHI and BglII sites of pBR322ΔRIBGL. Transformants with the correct orientation displayed fragments of approximately 2500- and 1600-bp upon digestion with EcoRV, and were called pAO851.

The approximately 3000-bp PvuII fragment from pAO804 [Example 1g], encoding the second half of the AOX1 terminator and the 3' HIS4 gene, was isolated and inserted into the PvuII site of pAO851. Transformants with the correct orientation displayed fragments of approximately 650- and 6500-bp upon digestion with HindIII and approximately 2000- and 5500-bp upon double-digestion with BamHI and SalI, and were called pAO852.

The approximately 3000-bp ClaI-SalI fragment from pAO815 (Example 1h), encoding the 5' AOX1 sequence, the AOX1 terminator, and the 3' end of the HIS4 gene, was isolated and inserted into the ClaI and SalI sites of pAO852. Transformants with the correct orientation were linearized with BglII and displayed fragments of approximately 3000- and 4700-bp upon double-digestion with ClaI and SalI. Positive transformants were called pAO855.

The approximately 2000-bp EcoRI-SalI fragment from pHIS103M, encoding the AOX1 terminator and the 3' end of HIS4, was isolated and inserted into the EcoRI and SalI sites of pAO855. The resulting vector pAO856 (see Figure 2) was linearized with EcoRV and displayed fragments of approximately 430- and 7300-bp upon digestion with HindIII.

Example 2

Development of HSA-secreting strains

Expression vectors pHSA111, pHSA211, pHSA212, pHSA214, and pHSA216 were used to develop Mut$^+$ strains of Pichia pastoris for the expression of HSA. The phenotype Mut$^+$ refers to methanol utilization normal. Transformation of host strain GS115, a his4$^-$ mutant of Pichia pastoris deposited at ATCC under #20864, was accomplished by the spheroplast method [Cregg, J.M., Barringer, K.J., Hessler, A.Y., and Madden, K.R. Mol. Cell. Biol. 5:3376 (1985); see also US 4,879,231]

The Mut$^+$ strains were generated by integration of the entire expression vector into the HIS4 locus by additive homologous recombination. For site-directed addition into the HIS4 locus, the expression vector was linearized by digestion with StuI, which cleaves the plasmid within the HIS4 region. In this additive integration, the AOX1 locus is undisturbed, and the transformants retain the Mut$^+$ phenotype of wild-type growth on methanol.

Characterization of the Strains by Southern Blot Analysis

Mut$^+$ transformants resulting from integration by addition of expression vectors pHSA111, pHSA211, pHSA212, pHSA214, and pHSA216 to the HIS4 locus of GS115 were initially screened for histidine prototrophy. DNA from His$^+$ transformants was analyzed by Southern blot hybridization to verify the site of integration of the plasmid and the number of copies that had integrated. Thus, chromosomal DNA was digested separately with BglII and StuI. Two sets of BglII digests were probed with pBR322-based plasmids containing either the 5' and 3' regions of the Pichia AOX1 gene, or the Pichia pastoris HIS4 gene. The StuI digests were probed with plasmid pMET-HSA (see Example I). In addition, chromosomal DNA from transformants resulting

from integration of plasmids pHSA214 and pHSA216 was double-digested with ClaI and BamHI and probed with the pBR322-based plasmid containing the AOX1 5' and 3' regions.

Based on the results of Southern blot analysis of DNA from strains generated by transformation of GS115 with the five types of expression vectors the following strains were chosen for further characterization:

Strain Name	Vector	Site of Integration	Vector Number	Cassette Number
G+HSA111S4	pHSA111	HIS4	1	1
G+HSA111S6	pHSA111	HIS4	1	1
G+HSA211S4	pHSA211	HIS4	1	1
G+HSA211S6	pHSA211	HIS4	1	1
G+HSA212S31	pHSA212	HIS4	1	2
G+HSA212S32	pHSA212	HIS4	1	2
G+HSA214S34	pHSA214	HIS4	1	4
G+HSA214S40	pHSA214	HIS4	1	4
G+HSA214S51	pHSA214	HIS4	1	4
G+HSA216S47	pHSA216	HIS4	1	6
G+HSA216S56	pHSA216	HIS4	1	6
G+HSA216S44	pHSA216	HIS4	1	6

Example 3

Fermentation of HSA strains

HSA-expressing strains of Pichia pastoris were grown in one-liter fermentations to evaluate and compare the long-term growth and HSA production characteristics of the strains. Because Pichia pastoris achieves such high cell densities during extended growth periods, the cell mass accounts for as much as 50% of the total volume of the fermentor at the conclusion of the fermentation. Therefore, one-liter fermentations of Pichia pastoris are conducted in two-liter fermentors and typically yield 700-1000 ml of cell-free broth.

The protocols for fermentations of HSA-secreting strains of Pichia pastoris consist of three separate phases:

1) growth on excess glycerol,

2) growth on limited glycerol, and

3) growth on limited methanol.

Cells are initially grown on glycerol in a batch mode. Because glycerol strongly represses the AOX1 promoter, the HSA gene, which is regulated by this promoter, is not expressed during this phase. Following exhaustion of the glycerol, a limited glycerol feed is initiated. Glycerol does not accumulate during this phase, but cell mass increases, and the AOX1 promoter is depressed. Finally, in the third phase, a methanol feed is initiated which fully induces the AOX1 promoter for the production of HSA. Three variations of this basic protocol were used to evaluate HSA-expressing strains. The variable that was altered in the three protocols was the amount of glycerol in the batch and fed-batch phases.

1. Low cell density fermentation protocol

Runs 861 & 869: G+HSA111S4 (1 HSA-HSA expression cassette)

Run 868: G+HSA111S6 (1 HSA-HSA expression cassette)

Run 865: G+HSA212S31 (2 AMF-HSA expression cassettes)

Run 866: G+HSA211S10 (1 AMF-HSA expression cassette)

The following describes a three-stage fermentation carried out with a 2% glycerol batch/40 ml/hr 50% glycerol fed batch, 6 ml/hr MeOH feed rate.

The fermentor was autoclaved with 500 ml 10X Basal Salts (52 ml/l 85% phosphoric acid, 1.8 g/l Calcium Sulphate-$2H_2O$, 28.6 g/l Potassium Sulfate, 23.4 g/l Magnesium Sulfate-7H2O, 6.5 g/l Potassium Hydroxide), 20 g glycerol and water added to a one-liter volume. After sterilization and cooling, 4 ml of YTM_4 trace salts (5.0

ml/l Sulfuric Acid, 65.0 g/l Ferrous Sulfate-$7H_2O$, 6.0 g/l Copper Sulfate-$5H_2O$, 20.0 g/l Zinc Sulfate-$7H_2O$, 3.0 g/l Manganese Sulfate-H_2O, 0.1 g/l Biotin) were added to the fermentation and the pH was adjusted to 5.0 with 50% ammonium hydroxide containing 0.2% Struktol J673 antifoam. The pH of the medium was maintained at 5 by addition of the same solution. Excessive foaming was controlled by the addition of a 5% solution of Struktol J673 solution. Temperature was maintained at 30°C and dissolved oxygen was maintained above 20% saturation by increasing agitation, concentration, aeration and supplementing with oxygen when needed. Inocula were prepared from cells grown overnight in buffered YNB (11.5 g/L KH_2PO_4, 2.66 g/L K_2HPO_4, 6.7 g/L yeast nitrogen base, pH 6) containing 2% glycerol. The fermentor was inoculated with 40-100 ml of the cultured cells which had grown to an OD_{600} of 1-8, and the batch growth regimen was continued for 18 to 24 hours. At the point of glycerol exhaustion, indicated by an increase in dissolved oxygen concentration, a glycerol feed (50% glycerol plus 12 ml/L PTM_1) was initiated at 10 ml/h. After four hours of glycerol feeding, the feed was terminated and a 100% methanol feed containing 12 ml/L PTM_1 was initiated at a feed rate of 2 ml/h. After three hours, the methanol feed was increased to 6 ml/h and maintained for greater than 72 hours of methanol.

2. Moderate cell density fermentation protocol

Run 876 and 877: G+HSA212S31 (2 AMF-HSA expression cassettes)

The following protocol describes a three-stage fermentation carried out with a 5% glycerol batch/100 ml 50% glycerol fed batch, 6 ml/h MeOH feed rate.

This protocol was identical to the low cell density protocol except that the amount of glycerol in the batch phase was increased from 20 to 50 grams and the glycerol feed rate was increased to 20 ml/h and run for 5 hours. Therefore, the total volume of glycerol feed added

to the fermentation was increased from 40 to 100 ml. The remainder of the fermentation protocol was followed as described above, except that the pH of Run 877 was maintained at 6.5 instead of 5.0.

3. High cell density fermentation protocol

Runs 890 & 891: G+HSA111S6 (1 HSA-HSA expression cassette)

Runs 878 & 879: G+HSA212S31 (2 AMF-HSA expression cassettes)

The following protocol describes a three-stage fermentation carried out with a 5% glycerol batch/200 ml50% glycerol fed batch, 6 ml/h MeOH feed rate.

This protocol differed from the previous protocols in that the initial batch phase contained 50 g glycerol, 500 ml 10X Basal Salts, and water added to a volume of 0.9 liters; reduced from a volume of one liter. The volume of glycerol feed added during the glycerol fed-batch phase was increased to 200 ml by starting with a feed rate of 20 ml/h and increasing to 30 ml/h after 2 hours, and then 35 ml/h after 2 more hours. After a total seven hours of glycerol feeding, (and 200 ml volume added), the glycerol feed was terminated and the methanol feed initiated at 2 ml/h and the protocol was followed as described above. The methanol feed rate in Run 879 was 9 ml/hr instead of 6 ml/hr.

Methods of Monitoring the Fermentations

The levels of the NH_4OH, antifoam, glycerol and methanol reservoirs were recorded at time points. The wet weight of the culture was also determined as an indicator of cell growth in the fermentor. For this purpose, several one-ml aliquots of the fermentor culture was centrifuged for four minutes in a microfuge, the supernatant was decanted, and the wet pellet was weighted. Methanol and ethanol concentrations in the supernatant were determined by gas chromatography using a PorapakQ column. In addition, four cell pellets of 175 mg each and five supernatants of 1 ml each were prepared from each 15-ml culture aliquot and

frozen for future analysis (i.e., immunoassay, electrophoretic analysis, and immunoblot analysis).

Fermentation Results

The data for the above-described fermentations are summarized below:

Run	Strain	Cassette Copy Number	Hours on MeOH	Cell Yield	HSA Secreted
861	G+HSA111S4	1	77	268	164
869	G+HSA111S4	1	76	305	149
868	G+HSA111S6	1	76	278	247
890	G+HSA111S6	1	89	306	335
891	G+HSA111S6	1	89	324	356
865	G+HSA212S31	2	76	271	764
866	G+HSA211S10	1	76	267	421
876	G+HSA212S31	2	89	329	1175
877	G+HSA212S31	2	89	301	528
878	G+HSA212S31	2	87	327	1418
879	G+HSA212S31	2	65	388	1401

EXAMPLE 4. CHARACTERIZATION OF FERMENTATION PRODUCTS

A. HSA ELISA Protocol

1. ELISA

a. Reagents

Authentic HSA standard and goat anti-HSA antibody conjugated to horseradish peroxidase were purchased from Organon Teknika Corporation (Durham, NC). Authentic HSA was purchased in lyophilized form from Organon Teknika Corporation (Durham, NC) and reconstituted with distilled water. The extinction coefficient at 280 nm of a 1% solution is 5.3 and was used to determine the concentration of the standard.

Goat anti-HSA antibody was obtained from Atlantic Antibodies (Scarborough, ME). The horseradish peroxidase substrate, o-phenylenediamine (OPD), was purchased from Sigma Chemicals (St. Louis, MO). All other

chemicals were reagent grade and obtained from general suppliers.

b. Method

The ELISA used for quantitation of HSA in cell-free fermentor broth is a double-antibody assay in which the HSA molecule is sandwiched between an anti-HSA antibody coated on a 96-well microliter plate and an anti-HSA antibody conjugated to horseradish peroxidase.

In a typical assay, 200 µl of goat anti-human HSA antibody diluted 1:500 with a carbonate coating buffer (15 mM Na_2CO_3; 35 mM $NaHCO_3$, pH 9.5) is added to each well of a 96-well microliter plate and incubated 60 minutes at 37°C. After three washes with Tris-buffered saline (TBST: 10 mM Tris-HCl, pH 7.5; 150 mM NaCl; 0.05% Tween-20), the plate is incubated with 200 µl of "Blotto" buffer (0.003% antifoam-A; 0.1% thimerasol; 5.0% nonfat dry milk; 1X PBS, pH 7.5; 0.05% Tween-20) overnight at 37°C to prevent nonspecific binding of subsequent reagents. The plate is washed as before, and aliquots of standard HSA or diluted unknown sample are added to the wells and incubated for 2 hours at 37°C. After the plate is washed as before, 200 µl of goat anti-human HSA antibody conjugated to horseradish peroxidase is added, and the plate is incubated for 2 hours at room temperature. The plate is then washed as before, and 200 µl of the substrate solution (10 mg of OPD in 0.0125% H_2O_2) is added. After a 15 minute incubation at room temperature in the dark, the reaction is terminated by the addition of 50 µl 4.5 N H_2SO_4 and the absorbance at 492 nm is measured.

The assay has a sensitivity of 0.1 ng and is linear to 1.0 ng.

2. SDS-PAGE

a. Sample preparation

Cell-free broth was examined by SDS-PAGE and immunoblotting. Cell-free broth samples were prepared by centrifuging fermentor cultures at 6500 x g for 10 minutes and decanting the broth. The broth was then diluted

two-fold with 2X Laemmli sample buffer (Nature 227:680, 1970) (0.125 M Tris-HCl, pH 6.8; 4% SDS; 20% glycerol; 0.005% bromophenol blue; 200 mM DTT) and boiled for 5 minutes. The sample buffer used for the reduced stained SDS-PAGE and immunoblot contained 20 mM and 200 mM dithiothreitol (DTT), respectively.

b. SDS-PAGE

SDS-PAGE analyses were performed essentially as described by Laemmli using a 10% polyacrylamide gel with a 4% polyacrylamide stacking gel. The electrophoreses were carried out on Mini-Protean gel apparatus (BioRad).

c. Coomassie Staining

Protein visualization by Coomassie staining was conducted by staining the SDS-PAGE gels for 30 minutes in 50% ethanol, 10% acetic acid, 5% TCA, and 200 mg/L Coomassie Brilliant Blue. The gels were rehydrated by incubation for 30 minutes in 10% ethanol, 10% acetic acid, 1% TCA, 50 mg/L Coomassie Brilliant Blue, and then destained in 10% ethanol and 10% acetic acid.

3. Characterization of HSA

Protein staining and immunoblot analysis of SDS-PAGE gels of reduced and non-reduced cell-free broth from fermentations revealed the presence of a predominant protein that migrated to the same position as authentic HSA (i.e., 65 kD non-reduced, 70 kD reduced). A smaller species (~30 kD non-reduced, ~32 kD reduced) also was seen in broth from some of the strains. An ~50 kD band was also detected in the reduced broth samples for all fermentations. A similar, if not identical, approximately 45 kD species has been detected in the broth of recombinant HSA produced in S. cerevisiae (Sleep, et al. Bio/Technology 8:42, 1990). A small band (~32 kD non-reduced, ~30 kD reduced) was detected in the broth from some of the strains.

A higher molecular weight species of >106 kD which is also methanol induction time dependent was detected only

in the non-reduced immunoblot of the strains, suggesting the formation of HSA multimeters through disulfides.

The invention has been described in detail with reference to particular embodiments thereof. It will be understood, however, that variations and modifications can be effected within the spirit and scope of the invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Davis, Geneva R
Provow, Sally P

(ii) TITLE OF INVENTION: Production of Human Serum Albumin in Methylotrophic Yeast Cells

(iii) NUMBER OF SEQUENCES: 26

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Fitch, Even, Tabin & Flannery
(B) STREET: 135 South LaSalle Street, Suite 900
(C) CITY: Chicago
(D) STATE: Illinois
(E) COUNTRY: USA
(F) ZIP: 60603

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US/07/650,040
(B) FILING DATE: 04-FEB-1991

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Seidman, Stephanie L.
(B) REGISTRATION NUMBER: 33,779
(C) REFERENCE/DOCKET NUMBER: 50857PCT

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (619)552-1311
(B) TELEFAX: (619)552-0095

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1758 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1758
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
GAC GCT CAC AAG TCT GAA GTC GCT CAC AGA TTC AAG GAT CTA GGT GAA      48
Asp Ala His Lys Ser Glu Val Ala His Arg Phe Lys Asp Leu Gly Glu
 1               5                  10                  15

GAA AAC TTC AAG GCT TTG GTT TTG ATT GCT TTC GCT CAA TAC TTG CAA      96
Glu Asn Phe Lys Ala Leu Val Leu Ile Ala Phe Ala Gln Tyr Leu Gln
            20                  25                  30

CAA TGT CCA TTC GAA GAC CAC GTC AAG TTG GTC AAC GAA GTT ACT GAA     144
Gln Cys Pro Phe Glu Asp His Val Lys Leu Val Asn Glu Val Thr Glu
        35                  40                  45

TTT GCT AAG ACC TGT GTT GCT GAC GAA TCT GCT GAA AAC TGT GAC AAG     192
Phe Ala Lys Thr Cys Val Ala Asp Glu Ser Ala Glu Asn Cys Asp Lys
    50                  55                  60

TCC TTG CAC ACT TTG TTC GGT GAC AAG TTG TGT ACT GTT GCT ACT TTG     240
Ser Leu His Thr Leu Phe Gly Asp Lys Leu Cys Thr Val Ala Thr Leu
65                  70                  75                  80

AGA GAA ACT TAC GGT GAA ATG GCT GAC TGT TGT GCT AAA CAG GAA CCA     288
Arg Glu Thr Tyr Gly Glu Met Ala Asp Cys Cys Ala Lys Gln Glu Pro
                85                  90                  95

GAA AGA AAC GAA TGT TTC TTA CAA CAC AAG GAC GAC AAC CCA AAC TTG     336
Glu Arg Asn Glu Cys Phe Leu Gln His Lys Asp Asp Asn Pro Asn Leu
            100                 105                 110

CCA AGA TTG GTT AGA CCA GAA GTC GAC GTT ATG TGT ACT GCT TTC CAC     384
Pro Arg Leu Val Arg Pro Glu Val Asp Val Met Cys Thr Ala Phe His
        115                 120                 125

GAC AAC GAA GAG ACT TTC TTG AAG AAG TAC TTG TAC GAA ATC GCC AGA     432
Asp Asn Glu Glu Thr Phe Leu Lys Lys Tyr Leu Tyr Glu Ile Ala Arg
    130                 135                 140

AGA CAC CCA TAC TTC TAC GCT CCA GAA TTG TTG TTC TTC GCT AAG AGA     480
Arg His Pro Tyr Phe Tyr Ala Pro Glu Leu Leu Phe Phe Ala Lys Arg
145                 150                 155                 160

TAC AAG GCT GCT TTC ACT GAA TGT TGT CAA GCT GCC GAC AAG GCT GCT     528
Tyr Lys Ala Ala Phe Thr Glu Cys Cys Gln Ala Ala Asp Lys Ala Ala
                165                 170                 175

TGT TTG TTG CCA AAG TTG GAC GAA TTG AGA GAC GAA GGT AAG GCT TCT     576
Cys Leu Leu Pro Lys Leu Asp Glu Leu Arg Asp Glu Gly Lys Ala Ser
            180                 185                 190

TCC GCT AAG CAA AGA TTG AAG TGT GCT TCC TTG CAA AAG TTC GGT GAA     624
Ser Ala Lys Gln Arg Leu Lys Cys Ala Ser Leu Gln Lys Phe Gly Glu
        195                 200                 205

AGA GCC TTC AAG GCC TGG GCT GTT GCT AGA TTG TCT CAA AGA TTC CCA     672
Arg Ala Phe Lys Ala Trp Ala Val Ala Arg Leu Ser Gln Arg Phe Pro
    210                 215                 220

AAG GCT GAA TTT GCT GAA GTT TCT AAG TTG GTT ACT GAC TTG ACT AAG     720
Lys Ala Glu Phe Ala Glu Val Ser Lys Leu Val Thr Asp Leu Thr Lys
225                 230                 235                 240

GTT CAC ACT GAA TGT TGT CAC GGT GAC TTG TTG GAA TGT GCT GAC GAC     768
Val His Thr Glu Cys Cys His Gly Asp Leu Leu Glu Cys Ala Asp Asp
                245                 250                 255

AGA GCT GAC TTG GCT AAG TAT ATC TGT GAA AAC CAA GAC TCT ATC TCT     816
Arg Ala Asp Leu Ala Lys Tyr Ile Cys Glu Asn Gln Asp Ser Ile Ser
            260                 265                 270

TCT AAG TTG AAG GAA TGT TGT GAA AAG CCA TTG TTG GAA AAG TCT CAC     864
Ser Lys Leu Lys Glu Cys Cys Glu Lys Pro Leu Leu Glu Lys Ser His
        275                 280                 285

TGT ATC GCT GAA GTT GAA AAC GAC GAA ATG CCA GCT GAC TTG CCA TCT     912
Cys Ile Ala Glu Val Glu Asn Asp Glu Met Pro Ala Asp Leu Pro Ser
    290                 295                 300
```

```
TTG GCT GCT GAC TTC GTT GAA TCT AAG GAC GTT TGT AAG AAC TAC GCT      960
Leu Ala Ala Asp Phe Val Glu Ser Lys Asp Val Cys Lys Asn Tyr Ala
305                 310                 315                 320

GAA GCT AAG GAC GTT TTC TTG GGT ATG TTC TTG TAC GAA TAC GCT AGA     1008
Glu Ala Lys Asp Val Phe Leu Gly Met Phe Leu Tyr Glu Tyr Ala Arg
                325                 330                 335

AGA CAC CCA GAC TAC TCC GTT GTT TTG TTG TTG AGA TTG GCT AAG ACT     1056
Arg His Pro Asp Tyr Ser Val Val Leu Leu Leu Arg Leu Ala Lys Thr
            340                 345                 350

TAC GAA ACT ACT TTG GAA AAG TGT TGT GCT GCT GCT GAC CCA CAC GAA     1104
Tyr Glu Thr Thr Leu Glu Lys Cys Cys Ala Ala Ala Asp Pro His Glu
        355                 360                 365

TGT TAC GCT AAG GTT TTC GAC GAA TTT AAG CCA TTG GTT GAA GAA CCA     1152
Cys Tyr Ala Lys Val Phe Asp Glu Phe Lys Pro Leu Val Glu Glu Pro
    370                 375                 380

CAA AAC TTG ATT AAG CAA AAC TGT GAA TTG TTC AAG CAA TTG GGT GAA     1200
Gln Asn Leu Ile Lys Gln Asn Cys Glu Leu Phe Lys Gln Leu Gly Glu
385                 390                 395                 400

TAC AAG TTC CAA AAC GCT TTG TTG GTT AGA TAC ACT AAG AAG GTT CCA     1248
Tyr Lys Phe Gln Asn Ala Leu Leu Val Arg Tyr Thr Lys Lys Val Pro
                405                 410                 415

CAA GTC TCC ACT CCA ACT TTG GTT GAA GTC TCT AGA AAC TTG GGT AAG     1296
Gln Val Ser Thr Pro Thr Leu Val Glu Val Ser Arg Asn Leu Gly Lys
            420                 425                 430

GTT GGT TCT AAG TGT TGT AAG CAC CCA GAA GCT AAG AGA ATG CCA TGT     1344
Val Gly Ser Lys Cys Cys Lys His Pro Glu Ala Lys Arg Met Pro Cys
        435                 440                 445

GCT GAA GAC TAC TTG TCT GTT GTT TTG AAC CAA TTA TGT GTT TTG CAC     1392
Ala Glu Asp Tyr Leu Ser Val Val Leu Asn Gln Leu Cys Val Leu His
    450                 455                 460

GAA AAG ACT CCA GTT TCT GAC AGA GTT ACT AAG TGT TGT ACT GAA TCT     1440
Glu Lys Thr Pro Val Ser Asp Arg Val Thr Lys Cys Cys Thr Glu Ser
465                 470                 475                 480

TTG GTT AAC AGA AGA CCA TGT TTC TCT GCC TTG GAA GTT GAC GAA ACT     1488
Leu Val Asn Arg Arg Pro Cys Phe Ser Ala Leu Glu Val Asp Glu Thr
                485                 490                 495

TAC GTC CCA AAG GAA TTT AAC GCT GAA ACT TTC ACT TTC CAC GCC GAC     1536
Tyr Val Pro Lys Glu Phe Asn Ala Glu Thr Phe Thr Phe His Ala Asp
            500                 505                 510
```

```
ATC TGT ACC TTG TCC GAA AAG GAA AGA CAA ATC AAG AAG CAA ACT GCT      1584
Ile Cys Thr Leu Ser Glu Lys Glu Arg Gln Ile Lys Lys Gln Thr Ala
        515                 520                 525

TTG GTT GAA TTG GTT AAG CAC AAG CCA AAG GCT ACT AAG GAA CAA TTG      1632
Leu Val Glu Leu Val Lys His Lys Pro Lys Ala Thr Lys Glu Gln Leu
    530                 535                 540

AAG GCT GTT ATG GAC GAC TTC GCT GCT TTC GTT GAA AAG TGT TGT AAG      1680
Lys Ala Val Met Asp Asp Phe Ala Ala Phe Val Glu Lys Cys Cys Lys
545                 550                 555                 560

GCT GAC GAC AAG GAA ACT TGT TTC GCT GAA GAA GGT AAG AAG TTG GTT      1728
Ala Asp Asp Lys Glu Thr Cys Phe Ala Glu Glu Gly Lys Lys Leu Val
                565                 570                 575

GCT GCT TCT CAA GCT GCT TTG GGT TTG TAA                              1758
Ala Ala Ser Gln Ala Ala Leu Gly Leu  .
            580                 585
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 255 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..255
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
ATG AGA TTT CCT TCA ATT TTT ACT GCA GTT TTA TTC GCA GCA TCC TCC       48
Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala Ser Ser
 1               5                  10                  15

GCA TTA GCT GCT CCA GTC AAC ACT ACA ACA GAA GAT GAA ACG GCA CAA       96
Ala Leu Ala Ala Pro Val Asn Thr Thr Thr Glu Asp Glu Thr Ala Gln
            20                  25                  30

ATT CCG GCT GAA GCT GTC ATC GGT TAC TCA GAT TTA GAA GGG GAT TTC      144
Ile Pro Ala Glu Ala Val Ile Gly Tyr Ser Asp Leu Glu Gly Asp Phe
        35                  40                  45

GAT GTT GCT GTT TTG CCA TTT TCC AAC AGC ACA AAT AAC GGG TTA TTG      192
Asp Val Ala Val Leu Pro Phe Ser Asn Ser Thr Asn Asn Gly Leu Leu
    50                  55                  60

TTT ATA AAT ACT ACT ATT GCC AGC ATT GCT GCT AAA GAA GAA GGG GTA      240
Phe Ile Asn Thr Thr Ile Ala Ser Ile Ala Ala Lys Glu Glu Gly Val
65                  70                  75                  80

TCT TTG GAT AAA AGA                                                 255
Ser Leu Asp Lys Arg
                85
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 72 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS

(B) LOCATION: 1..72
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
ATG AAG TGG GTT ACT TTC ATT TCT TTG TTG TTC TTG TTC TCT TCT GCT        48
Met Lys Trp Val Thr Phe Ile Ser Leu Leu Phe Leu Phe Ser Ser Ala
  1               5                  10                  15

TAC TCT AGA GGT GTT TTC AGA AGA                                      72
Tyr Ser Arg Gly Val Phe Arg Arg
            20
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 43 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
TGCTTTGGGT TTGTAAGAAT TCGGATCCCG TAATCATGGT CAT                      43
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
TTGTAAGAAT TCGGATCCCG TAAT                                           24
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
TGAAAGAGCC TTCAAAGCTT GGGCTGTTGC TAGATT                              36
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
CTTCAAAGCT TGGGCTG                                                   17
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GTATCTTTGG ATAAAAGAGA CGCTCACAAG TCTGAAGT 38

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GATAAAAGAG ACGCTCAC 18

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GCTTGCATGC CTGCAGAATT CATGAAGTGG GTTACTTTCA TTTCTTTGTT GTTCGACGCT 60

CACAAGTCT 69

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ACTTTCATTT CTTTGTTG 18

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

ATTTCTTTGT TGTTCTTGTT CTCTTCTGCT TACTCTAGAG GTGTTTTCAG AAGAGACGCT 60

CACAAGTCT 69

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TCTGCTTACT CTAGAGGT 18

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AATTCGATGA GATTTCCTTC AATTTTTACT GCA 33

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GTAAAAATTG AAGGAAATCT CATCG 25

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GAATTAATTC 10

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CAGCAGATCT GCTG 14

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GACGTTCGTT TGTGCGGATC CAATGCGGTA GTTTAT 36

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GGCCTCTTGC GGGATGTCCA TTCCGACAGC 30

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

TTGCGGGATG TCCATTCC 18

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

TTTGTGCAAG CTTATGGATC CGCTTTAATG CGGTAGT 37

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TTATGGATCC GCTTT 15

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

TTCGAGCTCG GTACCTAAGG ATCCTGAGAT AAATTTCA 38

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

TACCTAAGGA TCCTGAG 17

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

ACGTTGTCAC TGAAGGCGGC CGCAGTATCT ACAAACC 37

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

TGAAGGCGGC CGCAGT 16

THAT WHICH IS CLAIMED IS:

1. A DNA fragment comprising an expression cassette, wherein said expression cassette comprises, in the direction of transcription, the following DNA sequences:

(i) a promoter region of a methanol responsive gene of a methylotrophic yeast,

(ii) a DNA sequence encoding a polypeptide consisting of:

(a) a secretion signal sequence selected from:

(1) the S. cerevisiae AMF pre-pro sequence, including the processing site: lys-arg, or

(2) the native HSA signal sequence and

(b) an HSA peptide; and

(iii) a transcription terminator functional in a methylotrophic yeast, wherein said DNA sequences are operationally associated with one another for transcription of the sequences encoding said polypeptide.

2. A DNA fragment according to Claim 1 further comprising at least one selectable marker gene and a bacterial origin of replication.

3. A DNA fragment according to Claim 2 wherein said fragment is contained within a circular plasmid.

4. A DNA fragment according to Claim 1 wherein said sequence encoding an HSA peptide encodes the natural 585 amino acid form of HSA.

5. A DNA fragment according to Claim 1 wherein said methylotrophic yeast is a strain of Pichia pastoris.

6. A DNA fragment according to Claim 5 wherein said methanol responsive gene of a methylotrophic yeast and the transcription terminator are both derived from the P. pastoris AOX1 gene.

7. A DNA fragment according to Claim 6 further comprising 3'- and 5'-ends having sufficient homology with a target gene of a yeast host for said DNA fragment to effect site directed integration of said fragment into said target gene.

8. A DNA fragment according to Claim 1 further comprising 3'- and 5'-ends having sufficient homology with a target gene of a yeast host for said DNA fragment to effect site directed integration of said fragment into said target gene.

9. A DNA fragment according to Claim 1 containing multiple copies of said expression cassette.

10. A DNA fragment according to Claim 7 containing multiple copies of said expression cassette.

11. A DNA fragment according to Claim 9 wherein said multiple copies of said expression cassette are oriented in head-to-tail orientation.

12. A DNA fragment according to Claim 7, which is derived from a SalI digest of the Pichia expression vector pHSA111 or pHSA211.

13. A DNA fragment according to Claim 7, which is the Pichia expression vector pHSA211.

14. A DNA fragment according to Claim 10, which is derived from a SalI digest of the Pichia expression vector pHSA212, pHSA214 or pHSA216.

15. A DNA fragment according to Claim 10, which is the Pichia expression vector pHSA212, pHSA214 or pHSA216.

16. A methylotrophic yeast cell transformed with the DNA fragment of Claim 1.

17. A methylotrophic yeast cell according to claim 16 wherein said yeast is a strain of Pichia pastoris.

18. A methylotrophic yeast cell transformed with the DNA fragment of Claim 4.

19. A methylotrophic yeast cell according to Claim 18 wherein said yeast is a strain of Pichia pastoris.

20. A P. pastoris cell transformed with the DNA fragment of Claim 5.

21. A P. pastoris cell transformed with the DNA fragment of Claim 6.

22. A P. pastoris cell transformed with the DNA fragment of Claim 7.

23. A methylotrophic yeast cell transformed with the DNA fragment of Claim 8.

24. A methylotrophic yeast cell transformed with the DNA fragment of Claim 9.

25. A methylotrophic yeast according to Claim 24 wherein said yeast is a strain of Pichia pastoris.

26. A methylotrophic yeast cell transformed with the DNA fragment of Claim 11.

27. A P. pastoris cell transformed with the DNA fragment of Claim 12.

28. A P. pastoris cell according to Claim 27 wherein said cell is selected from strain G+HSA211S4 or G+HSA211S6.

29. A P. pastoris cell transformed with the DNA fragment of Claim 14.

30. A P. pastoris cell according to Claim 29 wherein said cell is selected from strain G+HSA212S31, G+HSA212S32, G+HSA214S34, G+HSA214S40, G+HSA214S51, G+HSA216S44, G+HSA216S47 or G+HSA216S56.

31. A culture of viable P. pastoris cells according to Claim 17.

32. A culture of viable P. pastoris cells according to Claim 28.

33. A culture of viable P. pastoris cells according to Claim 30.

34. A process for producing HSA, said process comprising growing the cells of Claim 16 under conditions allowing the expression of said expression cassette(s) in said cells, and the secretion of said HSA product into the culture medium.

35. A process according to Claim 34 wherein said methylotrophic yeast is a strain of Pichia pastoris.

36. A process according to Claim 34 wherein said cells are grown in a medium containing methanol as a carbon source.

37. A process according to Claim 34 wherein said cells have the Mut$^-$ phenotype.

38. A process according to Claim 34 wherein said cells have the Mut$^+$ phenotype.

39. A process according to Claim 38 wherein said cells are selected from strain G+HSA211S4, G+HSA211S6, G+HSA212S31, G+HSA212S32, G+HSA214S34, G+HSA214S40, G+HSA214S51, G+HSA216S44, G+HSA216S47 or G+HSA216S56.

40. As a composition of matter, the expression vector pAO856.

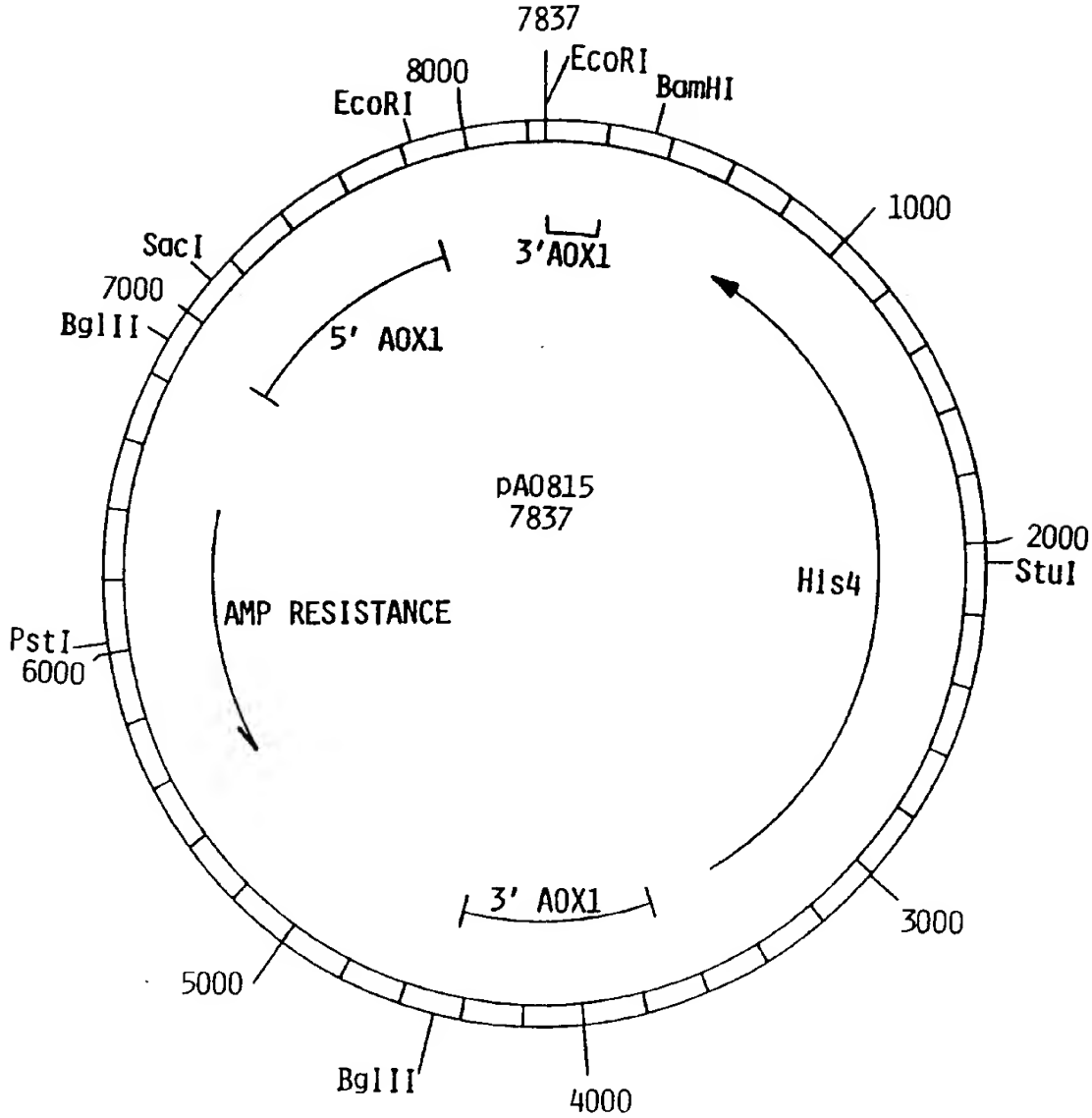
7837
EcoRI
BamHI
8000
EcoRI
1000
3'AOX1
SacI
7000
BglII
5' AOX1
pA0815
7837
2000
StuI
His4
AMP RESISTANCE
PstI
6000
3' AOX1
3000
5000
BglII
4000

FIG. 1

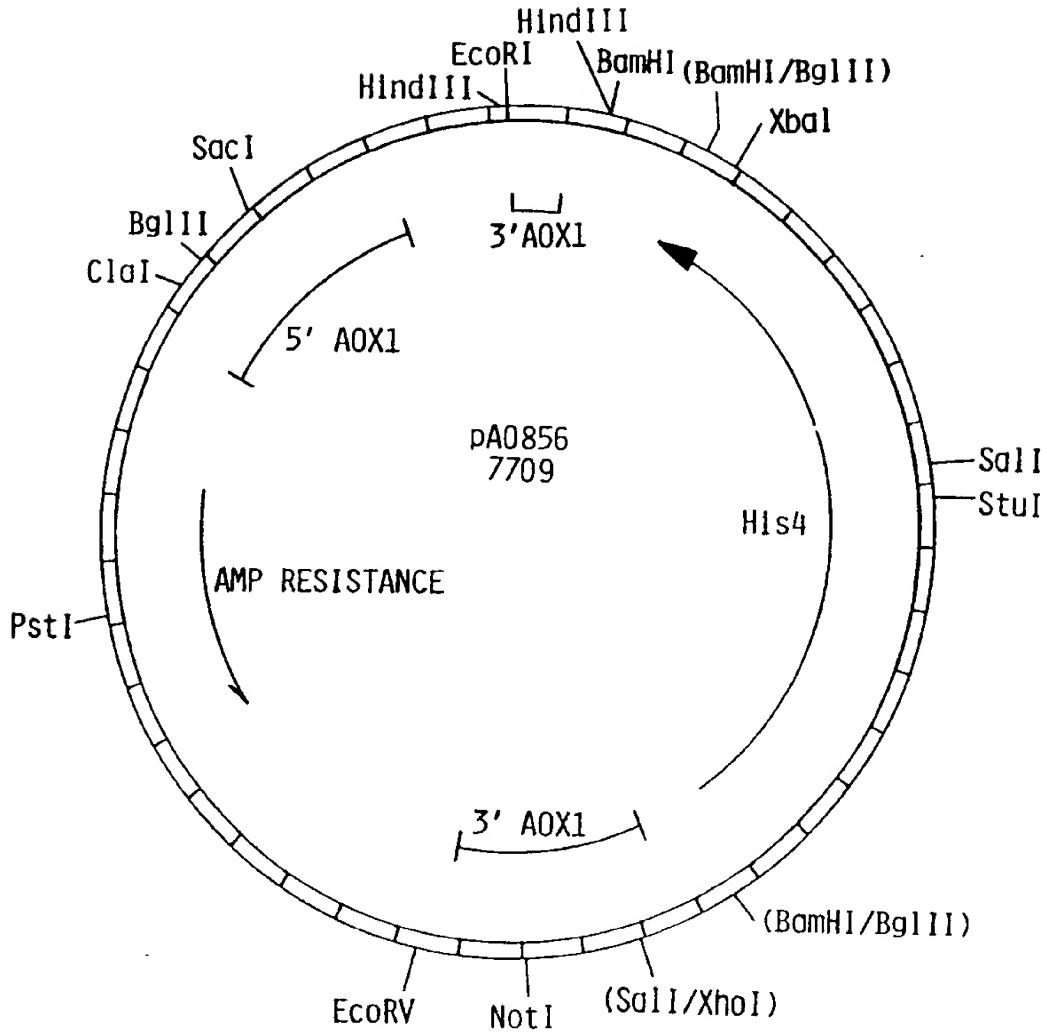
HindIII
EcoRI
BamHI
(BamHI/BglII)
HindIII
XbaI
SacI
BglII
ClaI
3'AOX1
5' AOX1
pA0856
7709
SalI
StuI
His4
AMP RESISTANCE
PstI
3' AOX1
(BamHI/BglII)
EcoRV
NotI
(SalI/XhoI)

FIG. 2

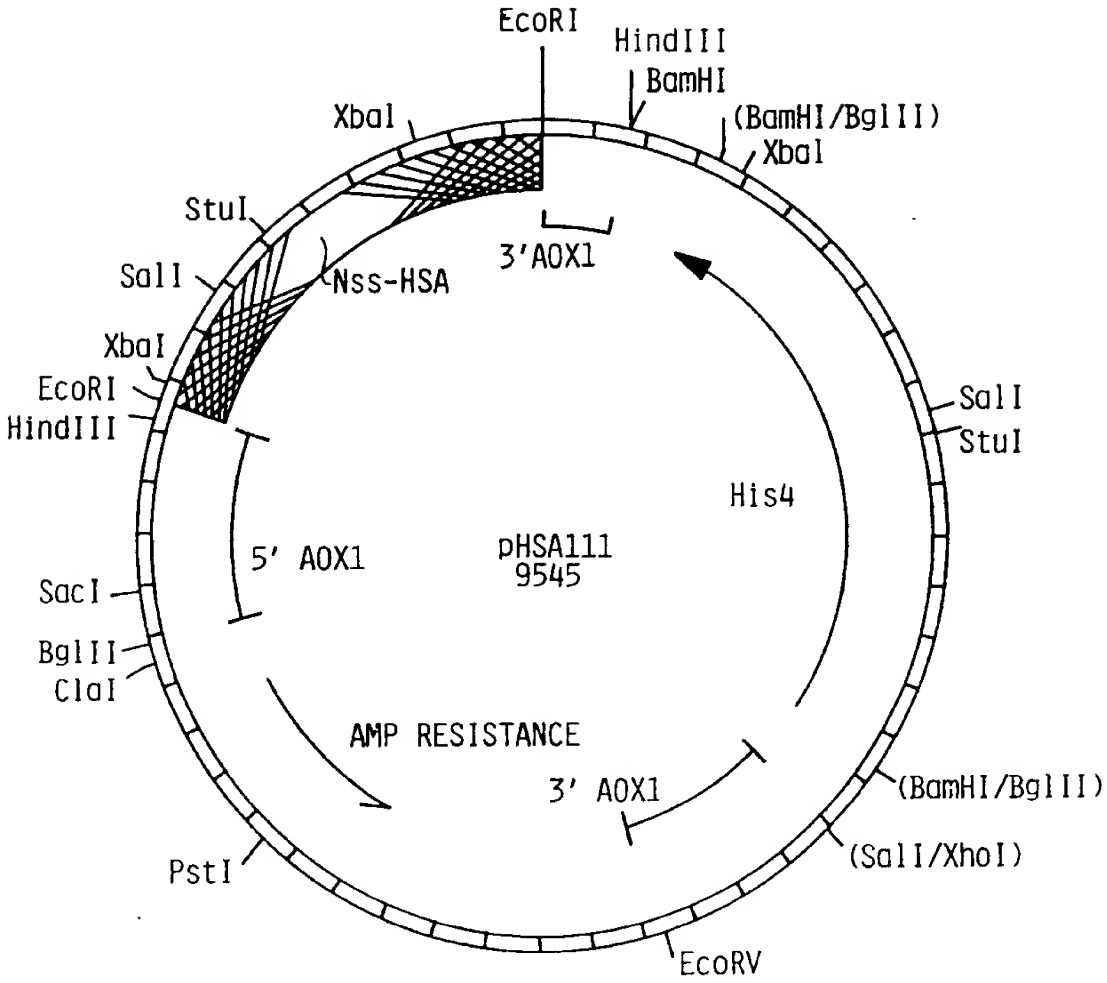
EcoRI
HindIII
BamHI
(BamHI/BglII)
XbaI
XbaI
StuI
3'AOX1
SalI
Nss-HSA
XbaI
EcoRI
HindIII
SalI
StuI
His4
5' AOX1
pHSA111
9545
SacI
BglII
ClaI
AMP RESISTANCE
3' AOX1
(BamHI/BglII)
(SalI/XhoI)
PstI
EcoRV

FIG. 3

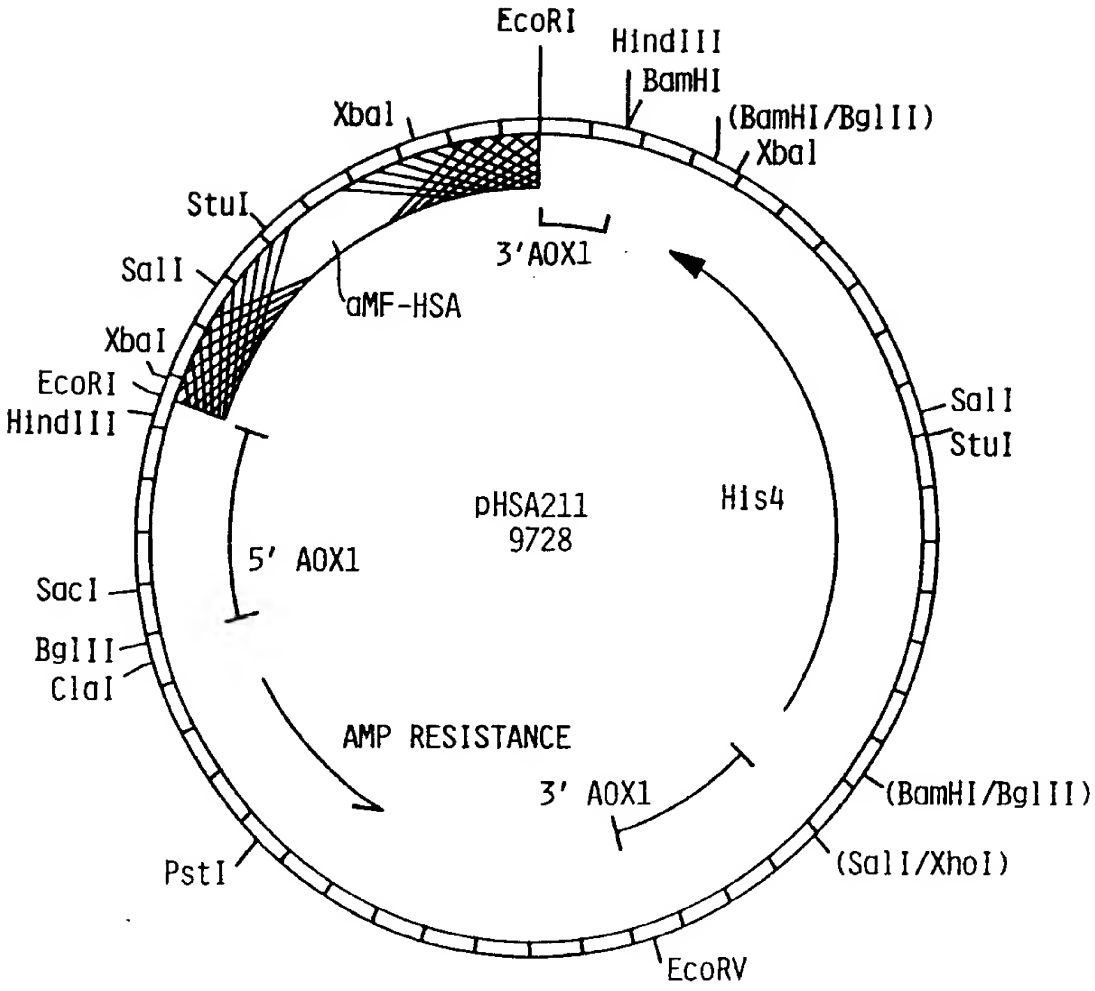
EcoRI
HindIII
BamHI
(BamHI/BglII)
XbaI
XbaI
StuI
SalI
3'AOX1
αMF-HSA
XbaI
EcoRI
HindIII
SalI
StuI
pHSA211
9728
His4
5' AOX1
SacI
BglII
ClaI
AMP RESISTANCE
3' AOX1
(BamHI/BglII)
(SalI/XhoI)
PstI
EcoRV

FIG. 4

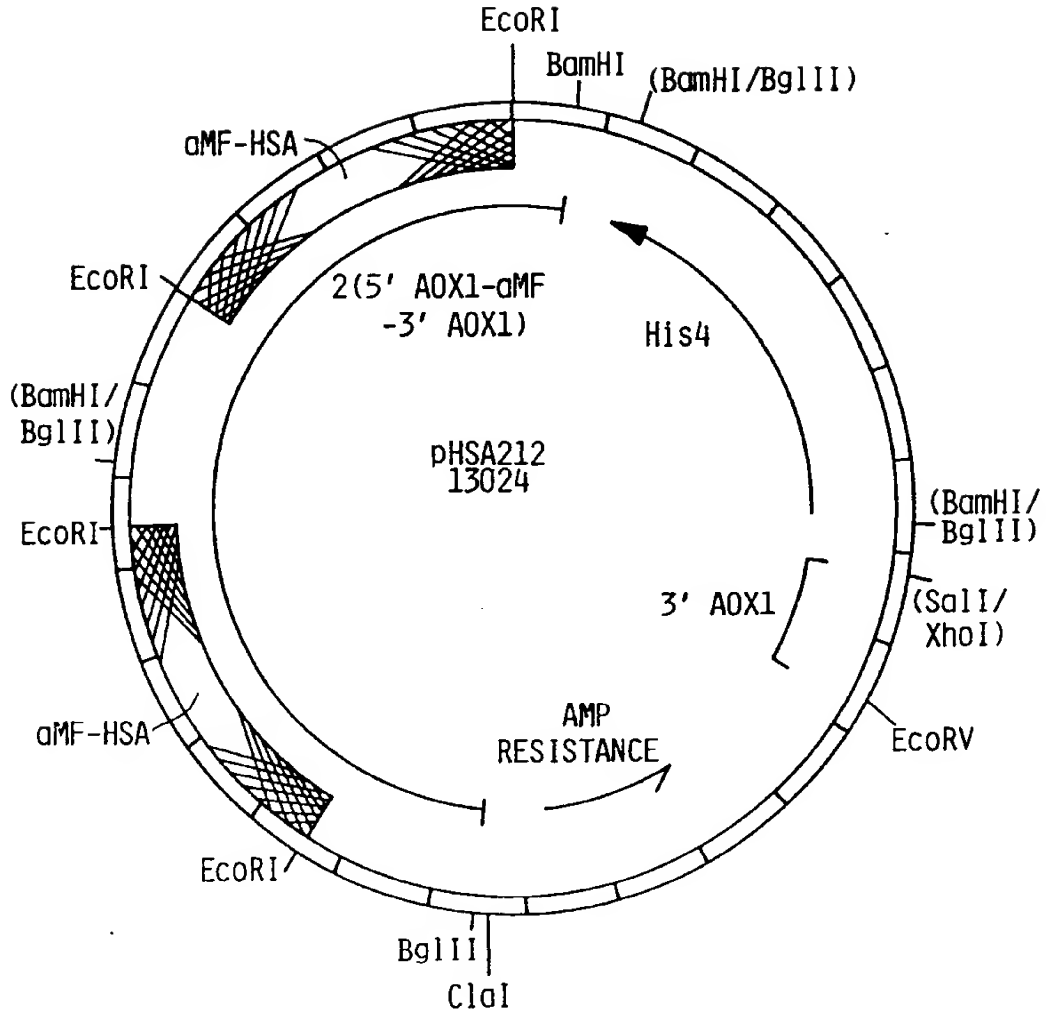
EcoRI
BamHI
(BamHI/BglII)
αMF-HSA
EcoRI
2(5' AOX1-αMF
-3' AOX1)
His4
(BamHI/
BglII)
pHSA212
13024
EcoRI
(BamHI/
BglII)
3' AOX1
(SalI/
XhoI)
αMF-HSA
AMP
RESISTANCE
EcoRV
EcoRI
BglII
ClaI

FIG. 5

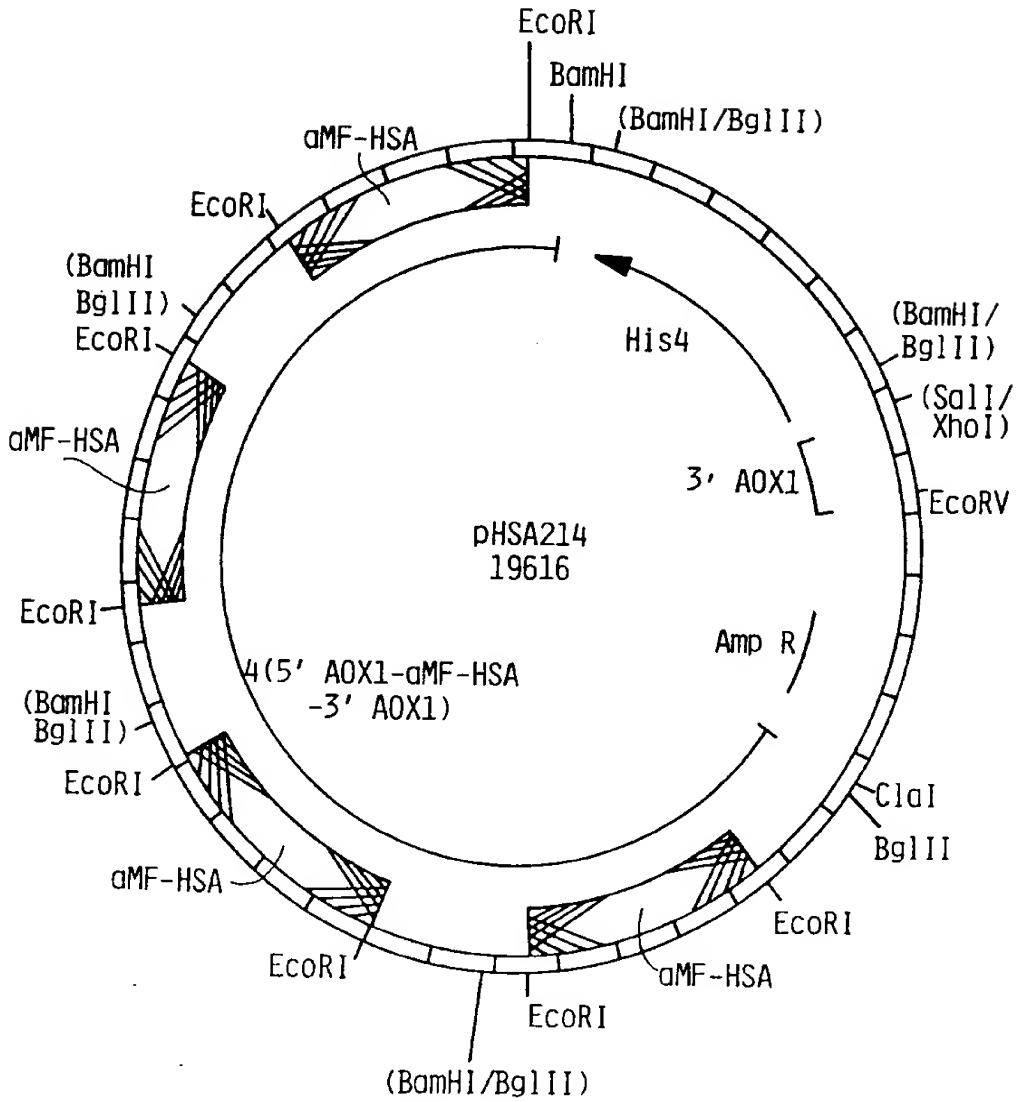
EcoRI
BamHI
(BamHI/BglII)
aMF-HSA
EcoRI
(BamHI
BglII)
EcoRI
aMF-HSA
EcoRI
(BamHI
BglII)
EcoRI
aMF-HSA
EcoRI
(BamHI/BglII)
His4
3' AOX1
pHSA214
19616
Amp R
4(5' AOX1-aMF-HSA
-3' AOX1)
(BamHI/
BglII)
(SalI/
XhoI)
EcoRV
ClaI
BglII
EcoRI
aMF-HSA
EcoRI

FIG. 6

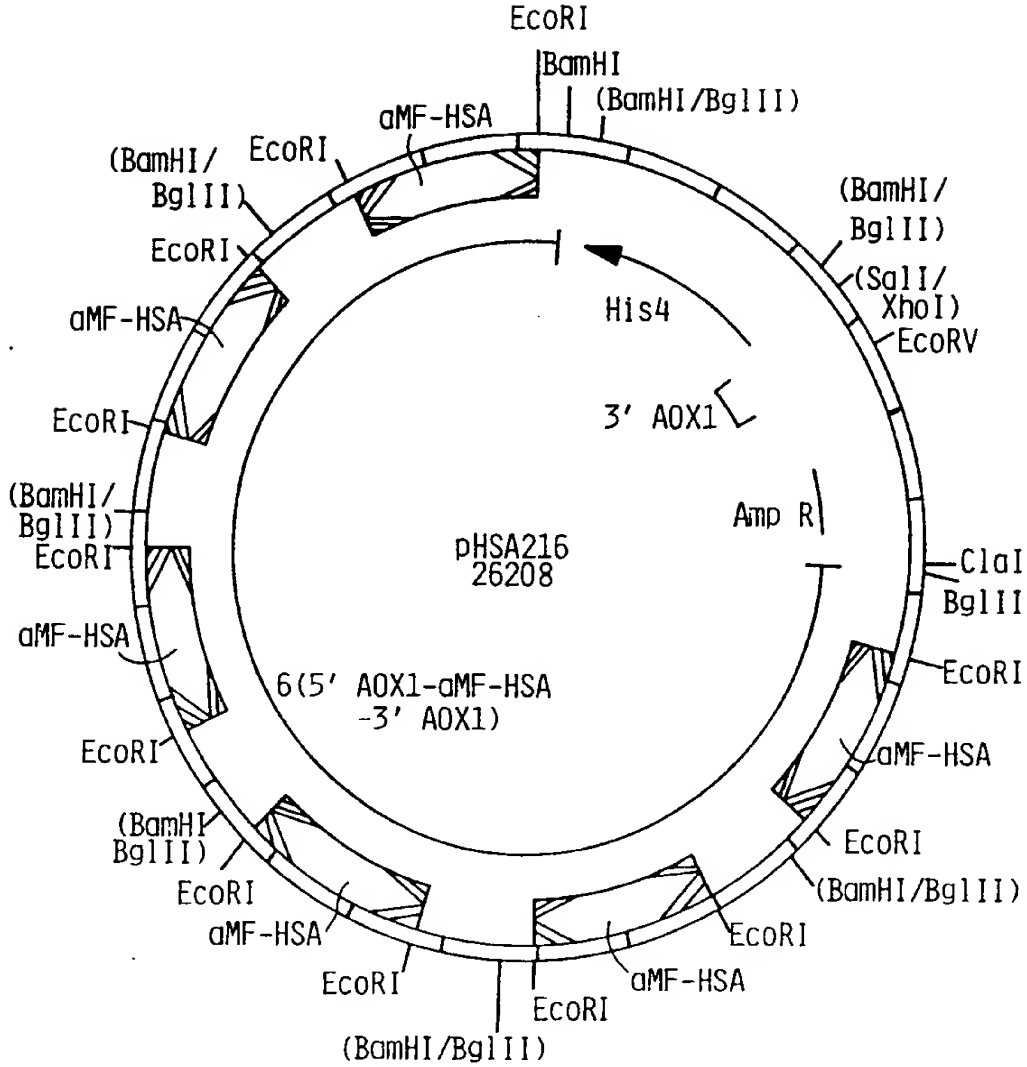
EcoRI
BamHI
(BamHI/BglII)
αMF-HSA
(BamHI/
BglII)
EcoRI
(BamHI/
BglII)
(SalI/
XhoI)
EcoRV
His4
3' AOX1
Amp R
pHSA216
26208
ClaI
BglII
EcoRI
αMF-HSA
EcoRI
(BamHI/BglII)
EcoRI
αMF-HSA
EcoRI
(BamHI/BglII)
EcoRI
αMF-HSA
EcoRI
(BamHI/
BglII)
EcoRI
αMF-HSA
EcoRI
(BamHI/
BglII)
EcoRI
αMF-HSA
6(5' AOX1-αMF-HSA
-3' AOX1)

FIG. 7

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US92/01015

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[3]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC (5): C12N 15/14, 15/81
US CL : 435/ 69.6, 69.8, 255, 320.1

II. FIELDS SEARCHED

Minimum Documentation Searched[4]

Classification System	Classification Symbols
U.S.	435/ 69.6, 69.8, 255, 320.1

Documentation Searched other than Minimum Documentation to the extent that such Documents are included in the Fields Searched[6]

III. DOCUMENTS CONSIDERED TO BE RELEVANT[14]

Category*	Citation of Document,[16] with indication, where appropriate, of the relevant passages[17]	Relevant to Claim No.[18]
T	US, A, 5,102,989 (Sitrin et al.) 07 April 1992, see entire document.	
T,P	US, A, 5,082,778 (Overbeeke et al.) 21 January 1992, see entire document.	
T,P	US, A, 5,041,236 (Carpenter et al.) 20 August 1991, see entire document.	
T,P	US, A, 5,032,516 (Cregg) 16 JULY 1991, see entire document.	
T,P	US, A, 5,026,828 (Yamazaki) 25 JUNE 1991, see entire document.	
T,P	US, A, 5,011,915 (Yamazaki) 30 APRIL 1991, see entire document.	
T,P	US, A, 5,004,688 (Craig et al.) 02 APRIL 1991, see entire documnet.	
Y	US, A, 4,963,483 (Ellis) 16 OCTOBER 1990, see entire document.	1-40
Y	US, A, 4,929,555 (Cregg et al.) 29 MAY 1990, see entire document.	1-40

* Special categories of cited documents:[15]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search[2]	Date of Mailing of this International Search Report[2]
23 APRIL 1992	06 MAY 1992
International Searching Authority[1]	**Signature of Authorized Officer**[20]
ISA/US	John L. Le GUYADER

Form PCT/ISA/210 (second sheet)(May 1986) ß

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

Y	US, A, 4,895,800 (Tschopp et al.) 23 January 1990, see entire document.	1-40
Y	US, A, 4,885,242 (Cregg) 05 December 1989, see entire document.	1-40
Y	US, A, 4,882,279 (Cregg) 21 November 1989, see entire document.	1-40
Y	US, A, 4,879,231 (Stroman et al.) 07 November 1989, see entire document.	1-40

V. ☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE[1]

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. ☐ Claim numbers _, because they relate to subject matter (1) not required to be searched by this Authority, namely:

2. ☐ Claim numbers _, because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out (1), specifically:

3. ☐ Claim numbers _, because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. ☐ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING[2]

This International Searching Authority found multiple inventions in this international application as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.

2. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4. ☐ As all searchable claims could be searched without effort justifying an additional fee, the International Search Authority did not invite payment of any additional fee.

Remark on protest

☐ The additional search fees were accompanied by applicant's protest.

☐ No protest accompanied the payment of additional search fees.

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category*	Citation of Document,[16] with indication, where appropriate, of the relevant passages[17]	Relevant to Claim No.[18]
Y	US, A, 4,857,467 (Sreekrishna et al.) 15 August 1989, see entire document.	1-40
Y	US, A, 4,855,231, (Stroman et al.) 08 August 1989, see entire document.	1-40
Y	US, A, 4,837,148 (Cregg))6 June 1989, see entire document.	1-40
Y	US, A, 4,818,700 (Cregg et al.) 04 April 1989, see entire document.	1-40
Y	US, A, 4,816,564 (Ellis et al.) 28 March 1989, see entire documnet.	1-40
Y	US, A, 4,812,405 (Lair et al.) 14 March 1989, see entire document.	1-40
Y	US, A, 4,808,537 (Stroman et al.) 28 February 1989, see entire document.	1-40
Y	Bio/Technology, Vol. 5, issued December 1987, Tschopp et al. "High-level secretion of glycosylated invertase in the methylotrohic yeast, Pichia pastoris", pages 1305-1308, see entire document.	1-40
Y	WO, A, 9003431 (Brierley et al.) 05 April 1990, see entire abstract, chemical abstract 189771t Vol. 113.	1-40
Y	WO, A, 9009434 (Salk Insitute) 23 August 1990, see entire abstract, abstract No. 90-14099.	1-40
Y	EP, A, 337492 (Craig et al.) 18 October 1989, see entire abstract, abstract No. C89-133643.	1-40
Y	Yeast, Vol. 6, Special Issue, s427, issued 1990, Romanos et al., "Expression of the tetanus toxin fragment C in Saccharomyces", see entire abstract, abstract No. 90-13480.	1-40
Y	WO, A, 9002810, (Salk Institute) 22 March 1990, see entire abstract, abstract No. 90-07612.	1-40
Y	ICSU Short Report, Vol. 8, issued 1988, Cregg et al., "High-level expression and secretion of heterologous proteins from the methylotrophic yeast Pichia pastoris", see entire abstract, abstract No. 89-11304.	1-40
Y	Dev. Ind. Microbiol., Vol. 29, issued 1988, Cregg et al., "Development of the methylotrophic reast, Pichia pastoris, as a host system for the production of foreign proteins", see entire abstract, abstract No. 89-00614.	1-40
Y	AU, A, 8814699, (Phillips Petroleum) 22 July 1988, see entire abstract, abstract No. 88-11111.	1-40
Y	GIM 90, Part I, issued 1990, Thill et al., "Positive and negative effects of mutiple copy intergrated expression vectors on protein expression in Pichia pastoris, pages 477-490, see entire document.	1-40
Y	WO, A, 9010697, (Salk Institute) 09 September 1990, see entire abstract, abstract No. 91-00231.	1-40

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category*	Citation of Document,[16] with indication, where appropriate, of the relevant passages [17]	Relevant to Claim No.[18]
Y	EP, A, 339567 (Phillips Petroleum) 02 November 1989, see entire abstract, abstract No. 90-01389.	1-40
Y	GB, A, 2217332 (Phillips Petroleum) 25 October 1989, see entire abstract, abstract No. 90-00765.	1-40
Y	EP, A, 263311, (Philips Petroleum) 13 April 1988, see entire abstract, abstrcat No. 88-06550.	1-40
X	EP, A, 344459, (Phillips Petroleum) 06 December 1989, see entire abstract, abstract No. 90-02645.	1-40
X	Yeast, Vol. 6, Special Issue, s447, issued 1990, Sreekrishna et al., "Expression of HSA in Pichia pastoris, see entire abstract.	1-40
X	Yeast, Vol. 6, Special Issue, s435, issued 1990, Hodgkins et al., "Secretion of HSA from Hansenula polymorpha" see entire abstract.	1-40
T	US, A, 5,102,789 (Seigel et al.) 07 April 1992, see entire document.	
T,P	US, A, 5,002,876 (Sreekrishna et al.) 26 March 1991, see entire docuemnt.	